

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**This Page Blank (uspto)**

[19]中华人民共和国专利局

[51]Int.Cl<sup>6</sup>

G06F 17/22

G06F 17/30



# [12] 发明专利申请公开说明书

[21] 申请号 95192149.5

[43]公开日 1997年2月26日

[11] 公开号 CN 1144004A

[22]申请日 95.11.21

[30]优先权

[32]94.11.21 [33]JP[31]286865/94

[32]95.3.30 [33]JP[31]73942/95

[86]国际申请 PCT/JP95/02371 95.11.21

[87]国际公布 WO96/16374 日 96.5.30

[85]进入国家阶段日期 96.9.18

[71]申请人 欧姆龙公司

地址 日本京都

[72]发明人 长冈宏 桑理圣二 林雅明

山下康成 松尾文人

[74]专利代理机构 柳沈知识产权律师事务所

代理人 马莹

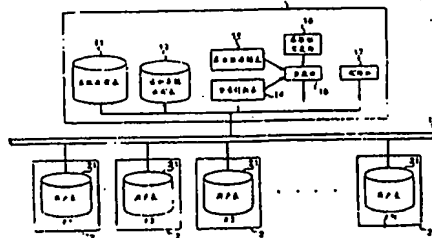
权利要求书 5 页 说明书 20 页 附图页数 27 页

[54]发明名称 多个客户机共有的数据库系统、数据更新方法及面向文字处理装置的应用

[57]摘要

在将多个用户数据库连接到主机数据库的数据库系统中，判断存储在各用户数据库的数据是否满足向主机数据库登录的条件，满足登录条件的数据将其记录在主机数据库，这种更新登录处理就所有的用户数据库的所有数据进行，把满足登录条件的所有数据登录到主机数据库，而登录的数据从用户数据库删除。

这种数据库系统也可以用到由日语、英语文书处理装置等通用辞典和用户辞典构成的系统中。



(BJ)第 1456 号

## 权 利 要 求 书

1、一种数据库装置，具有：主数据库；多个子数据库；把数据写入主数据库的写入装置；上述子数据库的数据分别具有特征量，以存储在上述子数据库的数据的特征量满足基准值为条件，通过上述写入装置把上述数据写入所述主数据库。

2、一种数据库装置，具有：主数据库；多个子数据库；把数据写入主数据库的写入装置；以同一个数据存储多个子数据库为条件、通过上述写入装置把上述数据写入所述主数据库。

3、一种数据库装置，具有：主数据库；与多个使用者一一对应的多个子数据库；把数据写入与上述使用者对应的子数据库的写入装置；用于输入检索条件的检索条件输入装置；根据上述检索条件检索上述主数据库或子数据库的检索装置；输出上述检索装置的检索结果的检索结果输出装置；基准值适合数据写入装置；

上述子数据库其各数据具有特征量，以存在上述数据的特征量满足基准值的基准值适合数据为条件，把上述适合基准值数据写入所述主数据库。

4、一种数据库装置，具有：主数据库；与多个使用者一一对应的多个子数据库；把数据写入与上述使用者对应的子数据库的写入装置；用于输入检索条件的检索条件输入装置；根据上述检索条件，检索上述主数据库或子数据库的检索装置；输出上述检索装置的检索结果的检索结果输出装置；及主数据写入装置，以相同的数据存储在规定数以上的子数据库为条件，把上述数据写入主数据库。

5、如权利要求3所述的数据库装置，其特征在于，以下述的基准值适合数据写入装置代替所述的基准值适合数据写入装置：即以多个子数据库中相同数据的特征量之和满足规定值为条件，把适合上述基准值数据写入所述主数据库。

6、如权利要求3所述的数据库装置，其中所述的基准值适合数据写入装置由以下的基准值适合数据写入装置替代，它以所述数据的特征量为基准值以上的基准值适合数据存在比规定数大为条件，将上述基准值适合数据写入主数据库。

7、如权利要求3所述的数据库装置，其中所述的基准值适合数据写入

装置由以下的基准值适合数据写入装置替代，它以所述数据的特征量在第一基准值以上的基准值适合数据存在比规定数大、并且上述特征量的和比第二基准值大为条件，将上述基准值适合数据写入主数据库。

5 8、如权利要求1、3、5、6、7中任意一项所述的数据库装置，其中，所述特征量是数据的使用频度。

9、如权利要求1、3、5、6、7中任意一项所述的数据库装置，其中，所述的特征量是数据的重要程度。

10 10、如权利要求4所述的数据库装置，其中，具有基准值适合数据删除装置，它以将上述基准值适合数据存于主数据库为条件，将上述基准值适合数据从子数据库中删除。

11、一种数据库系统，包括以下部分：

数据库服务器，它具有基准值适合数据写入装置，以存于主数据库、子数据库的数据的特征量满足基准值的基准值适合数据存在为条件，将上述基准值适合数据写入主数据库；

15 数据库终端，具有多个子数据库、写入装置、检索条件输入装置、数据库检索装置、检索结果输出装置，上述多个子数据库与多个用户一一对应，上述写入装置将数据写入与使用者对应的子数据库，上述检索条件输入装置用于输入检索条件，上述数据库检索装置根据上述检索条件检索上述主数据库或子数据库，上述检索结果输出装置输出从上述检索装置得到的检索结果。

20 12、一种数据库系统，包括以下部分：

数据库服务器，具有主数据库主数据写入装置，后者以相同的数据存于规定数以上的子数据库为条件，由主写入装置将上述数据写入主数据库；

25 数据库终端，具有多个子数据库、写入装置、检索条件输入装置、数据库检索装置、检索结果输出装置，上述多个子数据库与多个用户一一对应，上述写入装置将数据写入与上述用户对应的子数据库，上述检索条件输入装置用于输入检索条件，上述数据库检索装置根据上述检索条件检索上述主数据库或子数据库，上述检索结果输出装置输出从上述检索装置得到的检索结果。

13、一种数据库系统的数据更新方法，主数据库连接了多个可以交互通信的子数据库，其特征在于，

30 读出存于各子数据库中的各个数据；

判断上述读出的各个数据是否满足向预先设定的主数据库登录的条

件;

当断定登录条件满足时, 将该数据向主数据库追加。

14、如权利要求 13 所述的方法, 其中, 所述登录条件使用以下内容加以设定: 由用户给予的数据的特征量或重要程度、数据的使用频度、存储了  
5 同一数据的子数据库的数量是 1 或 2 以上。

15、如权利要求 13 所述的方法, 其中, 将向主数据库追加登录的数据从存有该数据的子数据库中删除。

16、一种文字处理装置, 包括具有各自的用户辞典的多个客户机、具有上述客户机共同使用的通用辞典的服务器, 其特征在于, 还具有以下部分:

10 基准值存储装置, 用于存储所选单词的基准值, 所选单词用于从登录于上述用户辞典的单词中向通用辞典登录;

使用频度记录装置, 记录登录于上述用户辞典的单词的使用频度;

15 登录装置, 判断登录于上述使用频度记录装置的用户辞典中登录的单词的使用频度是否超过了存储于上述基准值存储装置的基准值, 当上述单词的使用频度超过上述基准值时, 将该单词登录到上述通用辞典。

17、如权利要求 16 所述的文字处理装置, 其中, 具有单词删除装置, 在由上述登录装置将单词登录到上述通用辞典后, 将该单词从上述用户辞典中删除。

18、如权利要求 16 或 17 所述的文字处理装置, 其中, 上述基准值及上述单词的使用频度由登录单词的用户辞典数构成。  
20

19、如权利要求 16 或 17 所述的文字处理装置, 其中, 上述基准值及上述单词的使用频度由登录单词的用户辞典数及将该单词从上述用户辞典读出时的优先值之和构成。

20、如权利要求 16 至 19 任一项所述的文字处理装置, 其中, 具有基准  
25 值改变装置, 用于改变上述基准值存储装置的基准值。

21、如权利要求 16 至 19 任一项所述的文字处理装置, 其中, 上述通用辞典分为追加登录或不能删除的部分和追加登录或可以删除的部分。

22、一种文字处理装置, 具有客户机和服务器, 该服务器中具有仅供特定用户使用的用户辞典和供用户通用的通用辞典, 其特征在于包括: 基准值  
30 存储装置, 用于存储所选单词的基准值, 上述所选单词用于从登录于用户辞典的单词中向通用辞典登录;

单词使用频度记录装置，记录登录于上述用户辞典的单词的使用频度；  
登录装置，判断记录于上述单词使用频度记录装置的用户辞典所登录的单词的使用频度是否超过了存储于上述基准值存储装置的基准值，当上述单词的使用频度超过上述基准值时，将该单词登录到上述通用辞典；

- 5 单词删除装置，判断记录于上述单词使用频度记录装置的用户辞典所登录的单词的使用频度是否超过了规定的基准值，当使用频度在上述规定的基准值以下时，将该基准值以下的单词从用户辞典中删除。

- 23、如权利要求 22 所述的文字处理装置，其中，上述单词删除装置判断登录于上述单词使用频度记录装置的用户辞典中所登当的单词的使用频度是否超过了规定的基准值，当使用频度在上述规定的基准值以下时，将该基准值以下的单词从用户辞典中删除，而且，由上述登录装置把单词登录到上述通用辞典后，将该单词从上述用户辞典中删除。

24、如权利要求 22 或 23 所述的文字处理装置，其中，具有：

- 15 接续关系数据使用频度记录装置，用于存储具有单词和单词的接续关系的接续关系数据的使用频度；

接续关系数据删除装置，判断记录于该接续关系数据使用频度记录装置中的上述使用频度是否超过了规定的基准值，当上述使用频度在上述规定的基准值以下时，将上述基准值以下的接续关系数据从上述接续关系数据使用频度记录装置中删除。

- 20 25、如权利要求 23 或 24 所述的文字处理装置，其中，当处理的负荷在规定的基准以下时，自动地从上述用户辞典中将单词删除。

26、如权利要求 23 或 24 所述的文字处理装置，其中，当处理的负荷在规定的基准以下时，自动地从接续关系数据使用频度记录装置中将接续关系数据删除。

- 25 27、一种文字处理方法，具有多个客户机和服务器的文字处理装置中，所述多个客户机具有各自的用户辞典，所述服务器具有上述客户机通用的通用辞典，其特征在于，

- 30 存储选择单词的基准值，所选单词用于从登录于上述用户辞典的单词向上述通用辞典登录；记录登录于上述用户辞典的单词的使用频度；判断登录于上述用户辞典的单词的使用频度是否超过了上述基准值，在上述单词的使用频度超过上述基准值时，将该单词登录到上述通用辞典。

28、如权利要求 27 所述的文字处理方法，其中，将该单词登录到上述通用辞典后，从上述用户辞典中将该单词删除。

29、如权利要求 27 或 28 所述的文字处理装置，其中，上述基准值及上述单词的使用频度由登录单词的用户辞典数构成。

5 30、如权利要求 27 或 28 所述的文字处理方法，其中，上述基准值及上述单词的使用频度由登录了单词的用户辞典数及从上述用户辞典将该单词读出时的优先值之和构成。

31、一种文字处理方法，具有客户机及服务器的文字处理装置中，上述服务器具有仅供特定用户使用的用户辞典和供用户通用的通用辞典，其特征  
10 在于，

存储选择单词的基准值，该单词从登录于上述用户辞典的单词中登录到通用辞典；

记录登录于上述用户辞典的单词的使用频度；判断登录于用户辞典的单词的使用频度是否超过了存储的上述基准值，当判断上述单词的使用频度超  
15 过上述基准值时，将该单词登录到上述通用辞典；

判断登录于用户辞典的单词的使用频度是否超过了规定的基准值，当使用频度在上述规定的基准值以下时，将该基准值以下的单词从用户辞典中删除。

32、如权利要求 31 所述的文字处理方法，其中，判断登录于用户辞典的单词的使用频度是否超过了规定的基准值，当使用频度在上述规定的基准  
20 值以下时，将该基准值以下的单词从用户辞典中删除，而且，将单词登录到上述通用辞典后，将该单词从上述用户辞典中删除。

33、如权利要求 31 或 32 所述的文字处理方法，其中，包含单词与单词之间的接续关系的接续关系数据的使用频度被存储于文件中；

25 判断记录于该文件的上述使用频度是否超过了规定的基准值，在上述使用频度在上述规定的基准值以下时，将上述基准值以下的接续关系数据从上述文件中删除。

34、如权利要求 31 或 32 所述的文字处理方法，其中，当处理的负荷在规定的基准以下时，自动地从上述用户辞典中将单词删除。

30 35、如权利要求 33 所述的文字处理方法，其中，当处理的负荷在规定的基准以下时，自动地从存储有接续关系数据的使用频度的文件中将接续关系数据删除。

# 说明书

## 多个客户机共有的数据库系统、数 据更新方法及面向文字 处理装置的应用

5

本发明涉及通过网络将多个客户机连到共有数据库的数据库系统，更详细地涉及在各客户机的数据构成中可以将重复的数据有效地存于数据库的数据系统及数据库建立方法。另外，本申请还包括面向本发明的文字处理装置的应用。

10

众所周知，数据库系统是将多个客户机例如个人计算机通过网络连接到数据库装置而共享数据库。

15

在该种数据库系统中，共有数据库没有登录用户需要的数据时，在这种情况下，各用户利用所使用的客户机的存储器，生成各自的数据文件，供自己使用。

但是，用户生成各自的数据文件从系统整体考虑却有以下的低效率问题。

20

即，用户在各自的数据文件中重复拥有相同内容的数据，不仅在时间、人力上而且在资源上造成使用浪费，而且为了避免上面的使用浪费，在共有数据库中预先登录认为需要的所有数据会招致共有数据库的巨大化，导致数据利用率低下。

25

例如考虑如下翻译系统：服务器连接多个计算机终端，服务器共有作为数据库的翻译用辞典。这种情况，作为辞典，通过备齐各个专业领域的辞典才能进行各个领域的翻译，但是，这些辞典的所有内容都应用到的情况却很小，例如只限于一个专业领域的情况下，在预先准备的辞典中内容不全的问题却实际存在，实际上，谈到自然科学，还有生物、医学、化学、机械、电气、电子等多个领域，而且这些领域中的每一个还可细分为更多的领域，另外，通过研究的进展，细分化的一个领域本身也在变动。

因此，这种需求，从某种意义说是无限制的，服务器侧的辞典的扩张归根结底是无法对应的，之所以这样说，是因为由用户完备各自的辞典对用户来说是一种巨大的负担。

本发明的目的在于基本上消除共有数据库的数据库系统的上述问题。

为此，在本发明中，利用在各客户机独自生成的个别数据库来扩充服务器的数据库，即，服务器的数据库装置检索各用户的数据库(以下称用户数据库)中登录的数据，并判断是否应该登录到服务器的数据库。判断的基准既可以是对各用户登录的数据所授予的重要程度及特征量，也可以是对各登录数据的用户之间的重复度，或登录数据的使用频度，还可以是上述二者或三者的组合。

在此，所谓特征量和数据自身的信息是附加了特别意义的参数，例如使用者可以任意决定的数据库中的数据的重要程度，表现使用者如何频繁地存取数据库的数据等的数据使用频度，或者显示多少用户作为个人数据而存储在用户表中的登录数等。

无论哪种情况，用户数据库的数据中，只要是具有向服务器数据库登录价值的内容，便将此内容追加到服务器数据库中，并且将登录到服务器数据库的数据从用户数据库中删除，据此，用户数据库可以有效地利用自身的容量。

根据本发明，可以将用户需求高的数据有效地追加到服务器数据中，服务器数据库经充实可以共享有价值的数据库，因此，可以促进有效利用各用户侧的数据。

如上所述的服务器数据库的数据的登录处理未必是定期的、自动的，例如也可以通过操作人员的各自的判断，在需要时实行登录处理。

关于数据的登录条件(判断条件)，无需对全部数据千篇一律地设定登录条件，例如，数据按重复的种类划分的情况下可以按每个种类设定登录条件，而且比较理想的是该设定可以根据需要更改。而且，数据的登录条件所使用的数据的重要程度等有关数据登录的各参数由各用户给予，但因为是否可以登录的判断通常在服务器侧进行，所以，即使由用户将例如数据的重要程度(特征量)设定得很高，若不重复登录该数据，就不进行向服务器数据库的登录，可以保持多种适应性。

本发明可以应用到多种数据库系统，并可有效适用于各用户使用共同数据的可靠性高的系统，作为这样的系统，有例如文字处理系统(包括日语中的假名变换和翻译系统)、用户信息系统等。

在作为本发明应用的文字处理系统中，判断登录到用户辞典的单词的使

用频度是否超过了规定的基准值，当判断为单词的登录频度超过基准值的情况下，将该单词登录到追加系统辞典，并且，单词登录到追加系统辞典后，将该单词从用户辞典中删除。

5 而且判断登录在用户辞典的单词的使用频度是否超过了规定的基准值，当使用频度在规定的基准值以下时，将该基准值以下的单词从用户辞典中删除。

而且判断记录在文件中的接续关系数据的使用频度是否超过了规定的基准值，当使用频度在基准值以下时，将基准值以下的接续关系数据从文件中删除。

10 特别地，当处理的负荷在规定的基准以下时，自动地将单词从用户辞典中删除，而且将接续关系数据从文件中删除。

附图的简要说明：

图 1 为本发明的第一实施例的数据库系统的系统构成图；

图 2 是显示向第一实施例的系统共有表进行记录登录处理的流程图；

15 图 3 表示用户表一例的说明图；

图 4 表示将图 3 所示的用户表的记录进行登录处理后的状态说明图；

图 5 是本发明的第二实施例的数据库系统的系统构成图；

图 6 是图 5 所示的信息终端的构成方框图；

图 7 表示图 5 所示的主数据库文件的记录构成例的示意图；

20 图 8 表示图 6 所示的信息终端的辅助数据文件的构成例示意图；

图 9 表示构成图 5 所示的子数据库信息暂存部的例子示意图；

图 10 是登录处理后的主数据文件的示意图；

图 11 表示向子数据库信息暂存部进行记录登录处理的流程图；

25 图 12 表示从子数据库信息暂存部向主数据库文件进行登录处理的流程图；

图 13 表示将登录处理后的记录进行删除处理的流程图；

图 14 表示本发明的第三实施例的假名汉字变换装置的构成方框图；

图 15 表示图 14 中的登录计数表的构成示意图；

图 16 表示图 14 的基准值存储表的构成示意图；

30 图 17 是说明本实施例的假名的汉字变换装置的记录更新处理的流程图；

- 图 18 表示本实施例的假名汉字变换装置的登录处理的说明图;  
图 19 表示本发明的第 4 实施例的假名汉字变换装置的构成方框图;  
图 20 表示图 19 的用户机装置的构成方框图;  
图 21 表示图 19 的服务器装置的构成方框图;  
5 图 22 表示图 21 的频度文件的构成方框图;  
图 23 表示图 21 的频度文件的构成方框图;  
图 24 表示图 21 的关系频度文件的构成方框图;  
图 25 表示图 21 的关系频度文件的构成方框图;  
图 26 表示图 21 的项目变换表的构成方框图;  
10 图 27 表示将图 21 中的频度文件进行存取处理的方框图;  
图 28 表示本实施例的在线学习处理的概略流程图;  
图 29 表示本实施例的用户辞典进行单词删除处理的流程图;  
图 30 表示从用户辞典删除单词时进行评价处理的流程图;  
图 31 表示对用户辞典进行无用单元收集处理的流程图;  
15 图 32 表示将登录在本实施例的接续频度文件的关系频度进行降低处理的流程图;  
图 33 表示接续频度的推移图;  
图 34 表示将登录在本实施例的接续频度文件的接续关系数据进行删除处理的流程图;

20 下面根据附图说明本发明的数据库系统的实施例。

#### 〈第一实施例〉

图 1 表示本发明的数据库系统的第一实施例的构成方框图;

- 本实施例的数据库系统由连接在以太网 10 的服务器 1 和具有各自用户表 21 的 # 1 ~ # N 的客户机 2 构成, 其中客户机 2 由个人计计算机构成。服务器 1 由具有大容量存储器的计计算机构成, 从功能上看包括: 系统共有表 11、追加系统共有表 12、登录记数表 14、基准值存储表 15、登录部 16、记录删除部 17 及基准值变更部 18。
- 25

- 在此, 所谓服务器, 是指用 LAN 连接的计算机中提供信息和服务的工作站等计算机, 该服务器检索由客户机等输入的数据库的检索条件, 同时检索服务器的数据库, 将其检索结果提供给客户机。
- 30

而且客户机是从其他计算机接受服务的计算机, 表示终端和工作站等装

置。

在此，系统共有表 11 构成如下：它是通过所有的客户机 2 将共同使用的记录作为表构造的存储部，用户可以进行数据的追加和删除。

- 5 追加系统共有表 12 未登录在系统共有表 11 中，但根据本发明方法，它是由用于将认为最该登录的记录追加登录的存储部构成的表。登录计数表 14 是就某记录在 # 1 ~ # N 的所有用户表 21 中，记录该记录实际登录的用户表数(登录频度)的表，若位于一个用户表 21 中的记录被登录，与该记录对应的用户表数只更新 1。

- 10 基准值存储表 15 是从向 # 1 ~ # N 的所有用户表 21 分别登录的记录中选择用于向追加系统共有表 12 登录的记录的基准值存储表，例如，可以根据记录的种类和属性设定基准值。

登录部 16 的构成如下：判断存储在登录计数表 14 的登录在用户表 21 的记录的登录频度是否超过了存储在基准值存储表 15 的基准值，当记录的登录频度超过了基准值时，将该记录向追加系统共有表 12 中登录。

- 15 记录删除部 17 的构成如下：记录向追加系统共有表 12 登录后，将该记录从 # 1 ~ # N 的所有用户表 21 中删除。

基准值变更部 18 根据需要改变基准值存储表 15 中的各基准值。

下面参照图 2 说明本实施例的数据库系统的工作。

- 20 登录处理开始后，首先，# 1 的客户机 2 从服务器 1 接收用户表读入命令后(S1)，读出登录在 # 1 用户表 21 中的 1 个记录(S2)，并且对应该读出的记录，将登录计数表 14 中的对应登录频度加 1。

然后，通过将取出的记录作为关键字从 # 2 顺序检索到 # N，判断是否登录在其他的 # 2 ~ # N 的所有用户表 21 中，若已登录，将每次的登录频度增加(S3)。

- 25 这样，就一个记录处理结束后，登录部 16 判断记录在登录计数表 14 中的登录频度是否超过了存储在基准值存储表 15 中的基准值(S4)，当该记录的登录频度超过了对应的基准值时(S4 中的是)向追加系统共有表 12 中登录(S5)的同时，向记录删除部 17 送登录结束的指令。

- 30 记录删除部 17 从登录部 16 接受登录结束的指令后，从选择的记录登录的用户表 21 中将该记录删除(S6)。

上述记录的登录处理，是就最初登录在 # 1 的用户表 21 的全部记录反

复进行(S7), 接着, 象 # 3、# 4 那样, 一直按用户顺序进行到最后的用户表(# N), 最后的用户表的最后的记录处理结束时, 结束第一次的登录处理(S8).

图 3 及 4 显示了具体的例子.

5 如图 3 所示, 考虑多个用户将各用户的电话号码本作为用户表登录在各客户机 2 的情况, 这种情况的 1 个记录由人个姓名、公司名、公司的电话号码、住宅的电话号码构成, 象图中用网络表示的那样, 假想在两个用户表中登录了重复的记录, 这时, 假设在追加系统共有表 12 中没有登录这些重复的记录.

10 登录值(登录条件)设为 2 人以上将同一记录登录在各用户表中.

现在, 使用该登录值, 进行用图 2 说明的登录处理后, 取出用户 # 1 和用户 # 2 共同登录的“长冈宏”这一记录及用户 # 1 和用户 # N 共同登录的“田中一”这一记录, 如图 4 所示登录到追加系统共有表, 在用户 # 1、用户 # 2、用户 # N 的各表中, 登录的记录分别被删除.

15 在上述实施例 1 中, 登录记录数表 14、基准值存储表 15、基准值变更部 18 等未必需要, 进行登录处理时, 也可以设定基准值, 而且, 登录处理也可以按所定的时序自动进行, 也可以由操作员间歇进行. 而且, 数据库的数据结构, 并不限于上述表结构, 也可以是具有预先设定的数据结构.

#### 〈第二实施例〉

20 图 5 表示多个信息终端 40 # 1 - # 5 通过 LAN(局域网)连接在数据库服务器 30 的数据库系统. 数据库服务器 30 具有: 通过信息终端 40 共享的主数据文件 31; 可以从信息终端 40 将数据自由追加、变更的辅助数据库信息暂存部 32. 处理部 33 根据存储在存储部(ROM) 34 的各种程序进行以下描述的各种处理.

25 在此, 信息终端以移动方便的小型个人计算机为例, 至少由以下几部分构成: LCD(液晶显示器)等显示装置、键盘等输入装置、RAM、ROM、由硬盘驱动器构成的存储器、及以将命令输出到上述装置进行控制的 CPU(Central Processing unit)为中心的处理装置.

30 图 6 表示信息终端 40 的构成方框图. 信息终端 40 将用户需要的主数据库文件 31 的一部分拷贝并存储到备份主数据库文件 41 中, 将用户可以自由地进行变更、追加的数据存储于数据库文件 42 中.

由信息终端 40 的输入部构成的检索条件输入后, 该检索条件通过处理部 43 暂时存储在信息终端 40 的存储部 47 的同时, 通过连接器 44 将检索条件传送到数据库服务器 30 侧. 数据库服务器 30 的处理部 33 根据传送的检索条件检索主数据库文件 31, 将检索结果传送给信息终端 40 的处理部 43. 信息终端 40 的处理部 43 将传送的检索结果显示在显示部 46.

而且, 信息终端 40 可以通过其自身具有的备份主数据库文件 41 或子数据库文件 42 进行数据库的检索. 与使用数据库服务器 30 时相同, 从输入部 45 输入检索条件后, 备份主数据库文件 41 或子数据库文件 42 由处理部 43 检索, 将检索结果送显示部显示.

图 7 表示主数据库文件 30 的一例, 该例中, 各记录由个人姓名、公司电话号码、住宅电话号码、使用频度(次数)、追加数据构成, 关于追加数据后面叙述.

图 8 表示信息终端 1 - 3 的子数据库文件的结构. 子数据文件的记录由个人姓名、公司电话号码、住宅电话号码及使用频度(次数)构成.

图 9 表示数据库服务器 30 的子数据库信息暂存部 32 的数据结构. 该暂存部 32 的各记录由个人姓名、公司电话号码、住宅电话号码、各使用者名、各使用者的使用频度及频度合计构成.

信息终端 40 连接到数据库服务器 30 后, 数据库服务器 30 顺序读出信息终端 40 的子数据文件 42 的记录, 将使用频度在基准值(如 25)之上的记录存储在子数据库信息暂存部 32. 使用者名和使用频度, 只有该使用频度比存储在子数据库信息暂存部 42 的使用频度大的情况下, 才更新该记录.

图 10 表示更新后的主数据库文件 31 的一例. 主数据库文件 31 的记录由个人姓名、公司电话号码、住宅电话号码、使用频度、追加数据构成, 这在前面叙述过, 与图 7 比较后可知, 追加了“冈本太郎”的记录, 追加数据栏中记录“1”, 即, 该追加数据栏用于判断该记录是在主数据库文件 31 中开始就有的呢, 还是新追加的.

这个追加记录条件设为例如在三个以上的信息终端 40、使用频度为 25 以上的记录, 参照图 9 的子数据库信息暂存部 32, 满足该条件的为“冈本太郎”这一记录. 因此, 如图 10 那样, 在登录处理(更新处理)中, “冈本太郎”这一记录被追加登录到主数据库文件 31, 则在追加数据栏中写入“1”. 在此, 将使用频度作为是否登录到主数据库文件 31 的基准(条件)之一; 但与使

用频度同时或代替使用频度，也可以将记录(数据)的重要程度作为基准。记录的重要程度在信息终端 40 的使用者在子数据库信息暂存部 32 作记录时由使用者的判断来决定。

5 图 11 表示读出信息终端 40 的子数据库文件 42 的记录，在子数据库信息暂存部 42 进行所需记录的追加更新的流程图。

判断某信息终端 40 是否连接到了数据库服务器 30，连接时(S40 是)，将该信息终端 40 的子数据库文件 42 的记录顺序读出(S41)，判断该记录的使用频度是否在规定值(如基准值 25)以上(S42)，若在规定值以上，将该记录存储到数据库服务器 30 的子数据库信息暂存部 32(S43)。该更新处理重复进行到连接的信息终端 40 的子数据库文件 42 的最后一个记录(S44)，最后的记录处理结束时，终止该程序，

图 12 表示向主数据库文件 31 进行记录写入操作的流程图。

15 该程序开始后，首先顺序读出记录在子数据库信息暂存部 30 的记录(S50)，判断该记录是否记录在规定数(如 3)以上的信息终端 40(S51)，是时，判断使用频度的合计值是否在规定值(在此，例如 80)以上(S52)，是时，因为满足追加登录条件，所以将该记录追加登录到主数据库文件 31 中(S53)。向主数据库文件 31 追加登录时，由于没必要在子数据库信息暂存部 32 保留该记录，所以将该记录从子数据库信息暂存部 32 中删除。

这种处理最好是一定时间如一日一次地进行，也可以不定期进行。

20 图 13 表示整理信息终端的子数据库文件的记录的流程图。因为将追加到主数据库文件 31 的记录再在信息终端 40 保存已无意义，所以将该记录删除。处理开始后，首先顺序读出主数据库文件 31 的记录(S60)，接着，读出信息终端 40 的子数据文件 42 的记录，判断子数据库文件 42 的记录和主数据库文件 31 的记录是否相同(S62)，是时，删除相同的子数据库文件 42 的记录(S63)，该删除处理是就子数据库文件 42 的所有记录进行，最后一个记录处理结束时，终止该删除处理(S64)，据此，可以有效利用子数据库文件 42 的存储容量。该删除处理经过一定时间间隔，并且以信息终端 40 连接到数据库服务器 30 为条件进行，但也可以由信息终端 40 的使用者在任意时间进行。

〈第三实施例〉

30 图 14 表示将本发明的数据库系统应用到作为日语文字处理装置的假名汉字变换装置时的构成方框图。

本实施例的假名汉字变换装置由作为连接到以太网 110 的服务器的服务器装置 100 和具有各自用户辞典 221 的 # 1 ~ # n 的客户机装置 200 构成，并且服务器装置 100 包括：系统辞典 111，追加系统辞典 112，使用频度信息文件 113，登录记数表 114，基准值存储表 115，登录部 116，单词删除部 117 及基准值变更部 118。

在此，系统辞典 111 是进行假名汉字变换时的共有的基本词辞典，对应一个读音指定几个变换语，这些指定的变换语设有词类及使用优先顺序。该系统辞典 111 不能追加和删除。

追加系统辞典 112 是进行假名汉字变换时将需要的变换语追加收集的单词辞典，登录有对应于读音的变换语，只是与系统词曲 111 不同，不设词类。

使用频度信息文件 113 存储有表示将对应于读音的变换语从系统辞典 111 及追加系统辞典 112 读出时的优先顺序的优先值。

登录记数表 114 是记录单词的登录频度的表，该单词的登录频度是就某记录登录在 # 1 ~ # n 的所有的用户辞典 221，如图 15 所示，由以下部分构成：每个固有名词、普通名词、动词等词类存储有单词读音的读音存储区 141，存储对应于该读音存储区 141 内的读音的变换语的变换语存储区 142，登录读音存储区 141 内的读音的用户辞典数存储区 143，存储读音存储区 141 内的读音的优先值之和的优先值之和存储区 144。

在此，所谓优先值，是指示将对应于某读音的变换语从用户辞典 221 中读出时的优先顺序的指标，由各用户将单词登录在用户辞典 221 时指定，而且该优先值之和是登录在各用户辞典 221 的同一单词的优先值之和。因此，优先值之和越大的读音的变换语，对于全体的用户辞典 221 而言，越优先读出。优先值之和便成为从 # 1 ~ # n 的所有用户辞典 221 登录的单词中选择向追加系统辞典 112 登录的单词的基准值。

在图中将存储在读音存储区 141 的读音“くぼ”进行变换后，为“窪”，登录该读音“くぼ”的用户辞典数为 10，优先值之和为 15。

基准值存储表 115 是存储选择单词的基准值的表，该单词用于从登录在 # 1 ~ # n 的所有用户辞典 221 的单词中向追加系统辞典 112 中登录，如图 16 所示，由以下几部分构成：存储每个词类名称的词类名称存储区 151，存储对应于该词类名称存储区 151 内的词类的第一基准值的第一基准值存储区 152，存储第二基准值的第二基准值存储区 153。

在此，第一基准值由登录对应某读音的变换语的用户辞典数 221 构成，第二基准值表示该优先值之和。

图中，固有名称的情况下，第一基准值是 3，第二基准值是 10。

5 登录部 116 判断登录在存储于登录记数表 114 中的用户辞典 221 中的单词的登录频度是否超过了存储于基准值存储表 115 中的基准值，单词的登录频度超过基准值的情况下，将该单词向追加系统辞典 112 中登录。

在此，所谓单词的登录频度是指用户辞典数，或用户辞典数与优先值之和的组合。

10 单词删除部 117 在单词向追加系统辞典 112 登录后，将该单词从 # 1 ~ # n 的所有用户辞典 221 中删除。

基准值变更部 118 根据需要可以改变基准值存储表 115 的第一基准值及第二基准值。

下面参照图 17 说明本实施例的假名汉字变换装置的工作。

15 每经过规定时间，# 1 客户机 200 从服务器 100 接受用户辞典读入命令后(S100)，读出登录在 # 1 的用户 221 中的一个记录(S110)，并且，将该读出的单词的读音、变换语、及优先值分别存储在登录计数表 114 的读音存储区 141、变换语存储区 142 及优先值之和存储区 144，同时将用户辞典数存储区 143 加 1。

20 然后，通过以取出的单词为关键字从 # 2 顺序检索到 # n，来判断是否登录在其他的 # 2 ~ # n 的所有用户辞典 221(S130)。

在已登录在 # 2 ~ # n 的用户辞典 221 的情况下(S130 是)，每次检索将用户辞典数存储区 143 的值加 1，而且将登录在该用户辞典 221 的优先值加到优先值之和存储区 144(S140)。

25 相反，在没有登录在其他的 # 2 ~ # n 的用户辞典 221 的情况下(S130 为 NO 否)，读出下一个单词并进行与上述相同的处理，这样，就登录在 # 1 的用户辞典的所有单词进行同样的处理后(S150)，从 # 2 的用户辞典 221 登录到 # n 的用户辞典 221 的单词不重复计算。

30 这样，读出 # n 的用户辞典 221 的最后的单词处理结束后(S160)，登录部 116 就记录在登录记数表 114 的所有读音，选择存储在用户辞典数存储区 143 的用户辞典数超过存储在基准值存储表 115 的第一基准值存储区 152 的第一基准值(用户辞典数)的部分(S170)，将该读音和与该读音对应的变换语登

录在追加系统辞典 112(S180), 同时向单词删除部 117 送登录结束的指令。

单词删除部 117 接受从登录部 16 送来的登录结束的指令后, 从选择的读音单词登录的用户辞典 221 中将该读音单词删除。

5 上述的 S170 的处理中, 登录部 116 也可以选择存储在优先值之和存储区 144 的优先值之和超过第二基准值的内容。

而且上述 S100 到 S160 的处理中, 接受从服务器 1 来的用户辞典读入命令时, 也可以使 # 1 ~ # n 的所有用户辞典 221 一起工作, 这样可以缩短选择时间。

参照图 18 具体说明上述的处理工作。

10 如图 18 所示, 在 # i 用户辞典 221 中, 与读音“ひろし”对应的变换语是“比吕志”, 它的优先值为 200, 在 # j 的用户辞典 221, 与读音“ひろし”对应的变换语为“比吕志”, 它的优先值为 20, 以这种情况为例说明。

这时, # i 的用户机 200 取出“ひろし”的变换语时, 登录在 # i 的用户频度信息 113 的优先值高的“宏”比优先值低的“博”优先选出, 关于 # j 15 的客户机 200 “博”比“宏”优先选出。

进行上述的处理后, 以读音“ひろし”将“比吕志”登录到追加系统辞典 112 后, 从 # i 的用户辞典 221 及 # j 的用户辞典 221 以读音“ひろし”的变换语“比吕志”被删除。

20 并且, 在 # i 的用户频度信息 113 中, 存储有“比吕志”的优先值, 因此, 将“比吕志”登录在追加系统辞典 112 后, “比吕志”最先选出。

在 # j 的使用频度信息 113 中, “比吕志”的优先值比“博”及“宏”的优先值低, 所以三者中最后选出。

在上面的实施中, 是就日语文字处理装置作的说明, 但本发明并不限于日语文字处理, 当然也可以应用到单词文字处理装置和机器翻译处理装置。

#### 25 <第四实施例>

图 19 表示本发明应用在假名汉字变换装置的第二实施例的构成方框图。

第二实施例的本假名汉字变换装置由连接在以太网 280 的 # 1 ~ # n 的客户机 250 和服务 260 构成。

30 # 1 ~ # n 的各客户机 250 如图 20 所示, 由以下几部分构成: 输入罗马字、假名等的键盘 251; 显示输入的文字、候补变换、变换结果等的显示

器 252；输入输出处理部 253；输入文字决定处理部 254。

在此，输入输出处理部 253 将从键盘 251 输入的文字串及将该文字串编辑的内容存储在缓冲区(未图示)，并在显示器 252 上显示。而且输入输出处理部 253 将对输入的文字串进行变换的结果存储在缓冲区(未图示)，并将该内容显示在显示器 252 上。

输入文字决定处理部 254 在从输入处理部 253 接受输入文字串后，通过后述的词典检索部 261，从系统词典 265、追加系统词典 266 及用户词典 267 得到该输入文字串的所有读音，并基于此对输入文字按文节分割。输入文字决定处理部 254 将按文节分割的输入文字按每个文节确定单词并将其输出到显示器 252。

另外，服务器 260 如图 21 所示，由以下几部分构成：词典检索部 261；数据记录处理部 262；离线学习处理部 263；登录语学习处理部 264；系统词典 265；追加系统词典 266；由 # 1 ~ # n 构成的用户词典 267；存储相对于读音使用的单词的使用频度的频度文件 268；存储词与词的接续关系数据的关系频度文件 269；项目变换表 270；上上次的频度文件 271；上次的频度文件 272；上上次的关系频度文件 273；上次的关系频度文件 274。

词典检索部 261 检索系统词典 265、追加系统词典 266 及用户词典 267，并检索对读音的候补变换。

数据记录处理部 262 在输入文字决定处理部 254 中，从客户机 250 接受决定单词候补变换的指令信息后，将把后述的频度文件 268 上的该单词的使用频度加 1 后的值作为新的使用频度更新，而且数据记录处理部 262 从客户机 250 接受使用了具有单词之间的关系的接续关系数据的指令信息后，将把关系频度文件 269 的关系频度加 1 后的值作为新的关系频度更新。

离线学习处理部 263 对词典检索部 261 发出锁命令，以禁止单词使用频度的更新、单词的登录、删除及词典的读入等，而且离线学习处理部 263 考查登录在用户词典 267 的单词的使用频度及接读关系频度、从用户词典 267 中删除使用频度显著降低(频度为 0)的单词。离线学习处理部 263 将因删除产生的用户词典 267 内的区域进行无用单元的收集，从而进行词典的再创建。

离线学习处理部 263 考查具有后述的关系频度文件 269 内的接续关系数据的关系频度，将关系频度在规定期间内无变化的接续关系数据的关系频度进行降低处理，在关系频度显著降低的 0 的情况下，从关系频度文件 269 中

将保持 0 关系频度的接续关系数据删除。

登录学习处理部 264 在登录在各用户辞典 267 的单词的使用频度超过规定的基准值的情况下，将超过的单词取出登录到追加系统辞典 266。

此时，登录学习处理部 264 将作为移动源的辞典 ID 的项目存储到后述  
5 的项目变换表 270。

系统辞典 265 是进行假名汉字变换时的单词辞典，具有只对应一个读音的单词，该单词中设置词类及使用优先顺序。系统辞典 265 可以将这些内容追加、删除。

而且，系统辞典 265 划分为法律、经济、科技、工程等领域，可以把划  
10 分的每个领域看作一个辞典。

这样一来，用户给所划分的每个领域设定辞典 ID，从中预先指定一个或多个，据此，只能检索该指定的辞典 ID 所在的领域，所以不必检索系统辞典 265 的整体，可以缩短变换时间。

追加系统辞典 266 与系统辞典 265 一样，是进行假名汉字变换时的单词  
15 辞典，对应一个读音指定几个变换语(汉字等)，在这些被指定的变换语中设有词类及使用优先顺序。

但是，追加系统辞典 266，象系统辞典 265 那样，没有区分法律、经济等领域，将使用优先顺序设定为 0。

用户辞典 267；根据自己的使用环境，用户在想使用新单词的情况下才  
20 有，是可以由用户自身进行单词登录的辞典，它的构成与系统辞典 265 一样，对一个读音指定几个变换语(汉字等)，在这些被指定的变换语中没有词类及使用优先顺序。

频度文件 266 如图 22 所示，由项目区 681 及频度区 682 构成。

在此，项目区 681 是存储在各辞典 ID 的单词(汉字等)中，就规定时间内  
25 读出的单词存储于该辞典 ID 中的项目。

频度区 682 表示存储在各辞典 ID 的单词(汉字等)中，在规定时间内对读音所使用的次数。

例如，图 23 所示，说明了在读音为“こっかい”的情况下频度文件 268 的数据存储状况。

30 如图所示，对应读音“こっかい”的单词(汉字)有“国会”及“黑海”两个，其中“国会”登记到作为政治领域的辞典 ID1 的第 123 号，表示在规

定期间内被读出 10 次。相反,“黑海”登记到作为地名领域辞典的辞典 ID2 的第 456 号,表示在规定时间内被读出 20 次。

关系频度文件 269 如图 24 所示,由以下部分构成:项目区 691;接续关系数据区 692,由接续地址辞典 ID 区 692a、项目区 692b 及关系频度区 692c 构成。

在此,项目区 691 表示象如“人”“去”那样,单词与单词连接的接续关系数据中,在前单词(如“人”)存储在辞典中的存储场所。

接续关系数据区 692 是接续关系数据中,表示在后语对在前语的接续关系的数据所存储的内容,在接续地址辞典 ID 区 692a,存储有保存后面单词的辞典 ID,在项目区 692b,存储有保存后面单词的辞典中的项目,在关系频度区 692c,存储在规定时间内上述变换文字被读出的次数。

如图 25 所示,以“人”“站立”、“人”“去”及“人”“出发”等例来说明。

如图所示,在“人”“站立”的情况下,“人”这个词被登记在作为一般辞典 1 的辞典 ID1 - 3 的第 20 号,“站立”这个词被登记在作为一般辞典的辞典 ID4 的第 30 号,表示与“人”的接续关系频度是 40。

“人”“去”的情况下,“人”这个词被登记在作为一般辞典的辞典 ID1 - 3 的第 20 号,“去”这个词被登记在作为一般辞典的辞典 ID4 的第 50 号,表示和“人”的接续关系频度为 60。

“人”“出发”的情况下,“人”这个词被登记在作为一般辞典的辞典 ID1 - 3 的第 20 号,“出发”这个词被登记在作为一般辞典的辞典 ID4 的第 60 号,表示与“人”的接续关系频度为 1。

项目变换表 270 从登录学习处理部 264 进行的用户辞典 267 向追加系统辞典 266 登录的情况下并在登录的单词存在于多个用户辞典 267 时,使这些登记项和追加系统辞典 266 的登记项对应。

用图 26 来说明,项目变换表 270 具有:表示从用户辞典 267 取出的单词的保存场所的追加系统辞典项目存储区 701,表示每个用户辞典 267 都有的、从追加系统辞典取出的单词的保存场所的用户辞典项目区 702,对于一个追加系统辞典项目存储区 701,对应有多个用户辞典项目区 702。

上上次的频度文件 271 与频度文件 268 的构成相同,是在离线学习处理部 263 从用户辞典 267 中将单词删除前,将上次频度文件 272 进行文件拷贝

形成。上次频度文件 272 与频度文件 268 及上上次频度文件 271 的构成相同，是在上上次频度文件 271 形成后，将频度文件 268 进行文件拷贝而形成。

上上次关系频度文件 273 的构成与关系频度文件 269 相同，是在上次频度文件 272 形成后，将上次关系频度文件 274 进行文件拷贝而形成。

- 5 上次关系频度文件 274 与关系频度文件 269 及上上次关系频度文件 273 的构成相同，是在上上次关系频度文件 273 形成后，将关系频度文件 269 进行文件拷贝而形成。

下面，说明本实施例的假名汉字变换装置的工作。

为了便于工作说明，分以下两种情况说明：〈 1 〉进行假名汉字变换的

- 10 情况 〈 2 〉进行将单词从用户辞典删除处理的离线学习的情况。

〈 1 〉反复进行假名汉字变换的情况

由于本假名汉字变换方法本身较熟悉，所以简单说明。

- 15 用户从键盘 251 输入用户 ID 后，从显示在显示器 252 上的一般领域 1、一般领域 2、法律领域、政治领域、经济领域、科技领域等中选择一个或多个领域后，通过输入输出处理部 253，将具有选择的领域的辞典 ID 和只有用户可以使用的用户辞典 267 的辞典 ID 输出到输入文字决定处理部 254(参照图 27)。

于是，输入文字决定处理部 254 对内置于服务器 260 的辞典检索部 261 发出指令，检索具有上述辞典 ID 的辞典及追加系统辞典(该辞典也有固有辞典 ID)。

20 然后，用户从键盘 251 输入输入文字串后，输入输出处理部 253 将从键盘 251 输入的输入文字串送显示器 252，在显示器 252 显示，同时将输入文字串作为输入文字输出到输入文字决定处理部 254。

- 25 输入文字决定处理部 254 接受输入文字后，利用关系频度文件 269，进行该输入文字的相关解析。

即，输入文字决定处理部 254 通过辞典检索部 261，判断输入文字是否存在于关系频度文件 269 中，在存在多个的情况下，将关系频度最高的句法结构(文章)作为与输入文字对应的文件输出。

- 30 另外，输入文字决定处理部 254 在输入文字不在关系频度文件 269 中的情况下，将输入文字按文节划分，对各文节决定规定的单词，从而确定与输入文字对应的文章。

若更详细地说明, 则输入文字决定处理部 254 从输入处理部 253 接受输入文字后, 为了将输入的文字按文节划分, 从系统辞典 265、追加系统辞典 266 及用户辞典 267 读出由输入及可以得到的所有读音, 并向辞典检索部 261 发出指令。

5 于是, 辞典检索部 261 从系统辞典 265、追加系统辞典 266 及多个用户辞典中只检索所指示的辞典 ID, 将候补读音的多个单词(同音异义词)输出到输入文字决定处理部 254。输入文字决定处理部 254 根据从辞典检索部 261 得到的读音, 进行输入文字的文节划分, 然后, 输入文字决定处理部 254 对每个文字进行单词变换。对每个文节所进行的单词变换如下方法进行。

10 即, 输入文字决定处理部 254 从辞典检索部 261 接受与读音对应的作为多个候补变换的单词后, 从频度文件 268 取出该单词的频度, 接着, 输入文字决定处理部 254 将单词的使用优先顺序和频度相加, 把值最大的确定为与读音对应的单词。

象上述那样输入文字被确定后, 数据记录处理部 262 在使用关系频度文件 269 确定输入文字的情况下, 把构成参与决定的接续关系数据的关系频度加 1 后的值作为新的关系频度, 另外, 在对每个构成输入文字的文节以选择特定单词来确定输入文字的情况下, 把将频度文件 269 中与上述特定的单词相当的频度值加 1 后作为新的频度值。

这样一来, 由于关系频度文件 269 的关系频度或频度文件 268 的频度值被一直更新下去, 所以可以使常被使用的输入文字或单词具有更高的变换率。

## (2)进行离线学习处理的情况

参照流程图来说明该离线学习处理。

图 28 表示离线学习处理的大概流程图。

25 系统负荷轻时如深夜, 若经过了预先设定的时间, 离线学习处理部 263 对辞典检索部 261 输出锁命令, 以使辞典检索部 261 不接受从输入文字决定处理部 254 来的检索申请(S151), 同时向登录学习处理部 264 输出该指令信息。

30 接着, 离线学心处理部 263 输出生成上上次频度文件 271、上次频度文件 272、上上次关系频度文件 273 及上次关系频度文件 274 的指令。

于是, 上上次频度文件 271 把上次的频度文件 272 进行文件拷贝而生

成, 然后, 上次频度文件 272 将频度文件 268 进行文件拷贝而生成, 接着, 上上次关系频度文件 273 把上次关系频度文件 274 进行文件拷贝而成生, 然后, 上次关系频度文件 274 将关系频度文件 269 进行文件拷贝而生成。接下来, 登录学习处理部 264 在登录在各用户辞典 267 的单词的使用频度超过规定值的情况下, 取出超过该基准值的单词向追加系统辞典 266 登录(S152)的同时, 将登录在追加系统辞典的单词从用户辞典 267 中删除。该处理详细记载在第三实施例中, 省略其说明。

接着, 离线学习处理部 263 将存储在 # 1 的用户辞典 267 的使用频度低的单词, 分别参照频度文件 268、上次频度文件 272 及上上次频度文件 273 的各个使用频度将其删除(S153), 该处理在后面叙述。

离线学习处理部 263 由于删除了单词, 在 # 1 的用户辞典 267 产生离散的空白区(碎片), 为了防止存储效果恶化, 进行再创建以收集无用单元, 消除不存在碎片数据的空白部分(S154), 该处理在后面叙述。# 1 的用户辞典 267 的再创建结束后, 离线学习处理部 263 参照上次关系频度文件 273、上上次关系频度文件 274 的各关系频度, 若需要, 将存储于关系频度文件的接续关系数据的关系频度降低(S155), 该处理在后面叙述。接着, 离线学习处理部 263 考查登录在接续关系频度文件 269 的接续关系数据的关系频度, 将关系频度为 0 的内容删除(S156), 该处理在后面详述。

接着, 离线学习处理部 263 判断是否存在下面应进行的用户辞典 267(S157), 因为存在下面应进行的 # 2 的用户辞典 267(S157 是), 离线学习处理部 263 返回步骤 S153, 关于 # 2 的用户辞典 267 进行同样的处理。

这样, 离线学习处理部 263 对从 # 1 的用户辞典 267 到 # n 的用户辞典 267 进行处理(S157 否)。

下面, 对概略描述离线学习处理的图 28 的流程图中的以下步骤进行说明: 进行单词删除处理的 S153, 进行用户辞典 267 的再创建处理的 S154, 降低关系频度 269 的关系频度的 S155 及删除关系频度 269 的接续关系数据的 S157。

(a)关于进行单词删除处理的步骤 S153

参照图 29 及图 30 所示的流程图来说明进行单词删除处理的步骤 S153。

离线学习处理部 263 读入用户辞典 267(S161), 就登录在所读入的用户辞典 267 的所有单词, 考查是否进行了评价(S162), 若判断为已就全部单词

进行了评价(S162 是), 结束处理。

相反, 或判断为还未就所有的单词进行评价(S162 否), 就重要单词的使用频度进行考查, 进行是否是应删除的单词的评价(S163)(下面, 将该处理叫做删除评价判断处理), 该删除评价判断处理用图 30 在后面叙述。

- 5        离线学习处理部 263 进行删除评价判断的结果, 重要的单词是应删除的单词的情况下(S163 是), 将该单词从用户辞典 267 中删除(S164), 相反不是应删除的单词的情况下(S163 否), 返回处理步骤 S162, 就登录的全部单词进行同样的处理。

在此, 参照图 30 说明上述的删除评价判断处理。

- 10       如图所示, 离线学习处理部 263 考查存储在频度文件 268 中的频度区 682(参照图 22)的频度值是否为 0(S171), 频度值不是 0 的情况下(S171 否), 断为重要的单词是不可以删除的单词(S177), 结束该单词的删除评价判断处理。

- 15       相反, 频度值为 0 的情况下(S171 是), 离线学习处理部 263 判断该单词是否登录在上次的频度文件 272 中(S172), 若未登录(S172 否), 判断为该单词是不可以删除的单词(S177), 结束该单词的删除评价判断处理; 若已登录(S172 是), 考查该单词的使用频度是否为 0(S173)。

- 20       离线学习处理部 263, 在该单词的使用频度不为 0 的情况下, 结束该单词的删除评价判断处理; 相反, 该单词的使用频度为 0 的情况下(S173 是), 判断该单词是否登录在上上次频度文件 271 中(S174)。

离线学习处理部 263 若判断没有登录时(S174 否), 判断为该单词是不可以删除的单词(S177), 结束该单词的删除评价判断处理; 相反, 若已登录(S174 是), 还要再考查该单词的使用频度是否为 0(S175)。

- 25       离线学习处理部 263 在该单词的使用频度不为 0 的情况下(S175 否), 判断为该单词是不可以删除的单词(S177), 结束该单词的删除评价判断处理; 相反, 该单词的使用频度为 0 的情况下(S175 是), 判断为该单词是可以删除的单词(S176), 结束删除评价判断处理。

(b)关于进行用户辞典 267 的再创建处理的步骤 S154

- 30       参照图 31 所示的流程图来说明进行用户辞典 267 的再创建处理的步骤 S154 的单词删除处理。

离线学习处理部 263 从用户辞典 267 的最初地址读取规定大小的区域

(S181), 判断该区域是否存在无用区(没有存储数据的离散的空白部分: 碎片)(S182).

5 离线学习处理部 263 在所读入的区域不存在无用区的情况下(S182 否), 返回处理步骤 S181, 进行同样的处理; 相反, 在存在无用区的情况下(S182 是), 删除该无用区并将该区以后的区域向前重新配置(无用单元收集), 再创建用户辞典 267(S183).

10 离线学习处理部 263 就所读入的区域把用户辞典 267 再创建后, 判断是否考查了用户辞典 267 的所有区域(S184), 未考查完所有区域的情况下(S184 否), 返回处理步骤 S181, 进行与上述同样的处理; 相反, 全部区域都已考查的情况下(S184 是), 结束处理.

(c)关于降低关系频度文件 269 的关系频度的步骤 S155

参照图 32 及图 33 所示的流程图来说明降低关系频度文件 269 的关系频度的步骤 S155.

15 离线学习处理部 263 判断是否考查了关系频度文件 269 的全部接续关系数据(S191), 当判断为已经全部考查的情况下(S191 是), 结束处理.

相反, 离线学习处理部 263 判断为还未全部考查完的情况下(S191 否), 则考查处于考查中的接续关系数据的关系频度(S193).

20 接着, 离线学习处理部 263 考查处于考查中的接续关系数据的关系频度是否存在于上次关系频度文件 273 中(S193), 若不在上次关系频度文件 274 的情况下(S193 否), 返回步骤 S192 进行同样的处理; 相反, 若存在于上次关系频度文件 274 中(S193 是), 就考查中的接续关系数据的关系频度考查是否存在于上上次关系频度文件 273 中(S194).

25 离线学习处理部 263 在不存在于上上次关系频度文件 273 的情况下(S194 否), 返回处理步骤 S192 进行同样的处理, 相反, 若存在于上上次关系频度文件 273 中(S194 是), 则将该接续关系数据的关系频度下降规定量后(S195), 返回处理步骤 S191 进行同样的处理.

利用图 33 更详细地说明上述处理.

图 33 表示关系频度的推移图.

如该图所示, 降低关系频度的情况, 有样板 1 及样板 2.

30 所谓样板 1 的情况, 是在上上次关系频度文件 273、上次关系频度文件 274 及关系频度文件 269 中, 关系频度没有发生变化的情况, 其理由是因为

目前为止该接续关系数据一次也没有使用。

所谓样板 2 的情况, 是过去的时间里接续关系数据的关系频度被降低, 之后关系频度没有变化的情况, 在该情况下, 与样板 1 的情况相同, 因为到目前为止, 该接续关系数据一次也没有使用。

5 (d)关于删除关系频度文件 269 的接续关系数据的步骤 S157

参照图 34 所示的流程图来说明删除关系频度文件 269 的接续关系数据的步骤 S157。

离线学习处理部 263 判断关系频度文件 269 的接续关系数据已被考查完 (S211), 若判断为已经全部考查的情况下 (S211 是), 结束处理。

10 相反, 离线学习处理部 263 若判断为还未全部考查完的情况下 (S211 否), 则考查处于考查中的接续关系数据的关系频度 (S212)。

接着, 离线学习处理部 263 考查处于考查中的该接续关系数据的关系频度是否为 0 (S213), 在不是 0 的情况下 (S213 否), 返回处理步骤 S212 进行同样的处理; 相反, 是 0 的情况下 (S213 是), 删除该接续关系数据 (S214)。根据上述发明, 判断登录在用户辞典的单词的登录频度是否超过了规定的基准值, 当判断为单词的登录频度超过基准值时, 因为将该单词向追加系统辞典登录, 无需使多个用户重新登录相同的单词, 所以能够使没有登录的用户第一次使用该单词时提高变换效率。

而且, 单词被登录到追加系统辞典后, 该单词被从用户辞典中删除, 所以可以节约用户辞典的数据区, 因此, 能够利用这些被节约的数据区进行其他处理。

判断登录在用户辞典的单词的使用频度是否超过了规定的基准值, 当使用频度在规定的基准值以下时, 因为这些基准值以下的单词被从用户辞典中删除, 所以把用户登录的单词中没被利用的单词从辞典中自动删除, 可以有效利用存储器资源的同时提高变换效率。

特别地, 当处理的负荷在规定的基准以下时, 单词被自动地从用户辞典中删除, 而且因为接续关系数据被从文件中删除, 所以可以有效利用存储器资源, 并且能提高变换效率。

图1

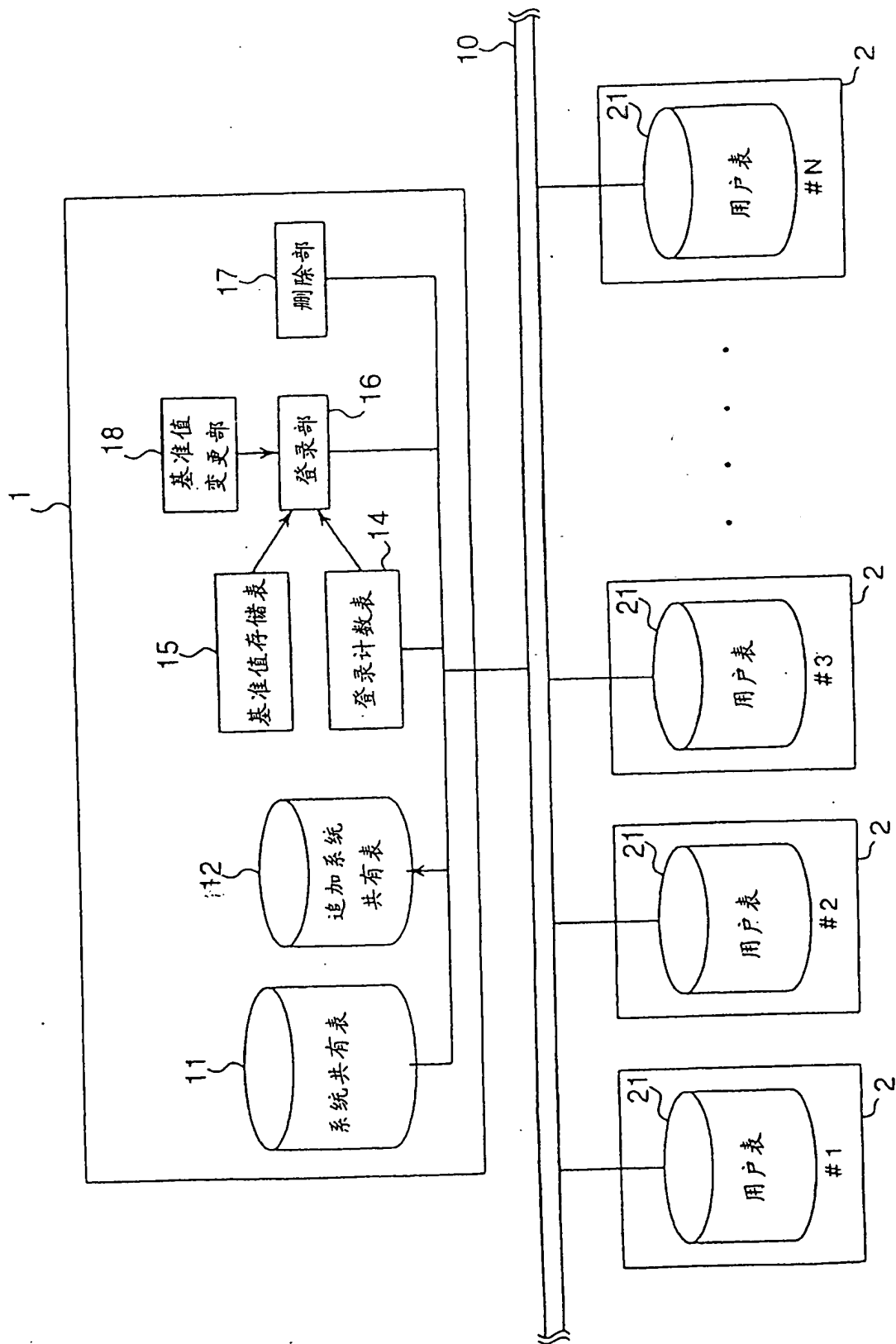


图 2

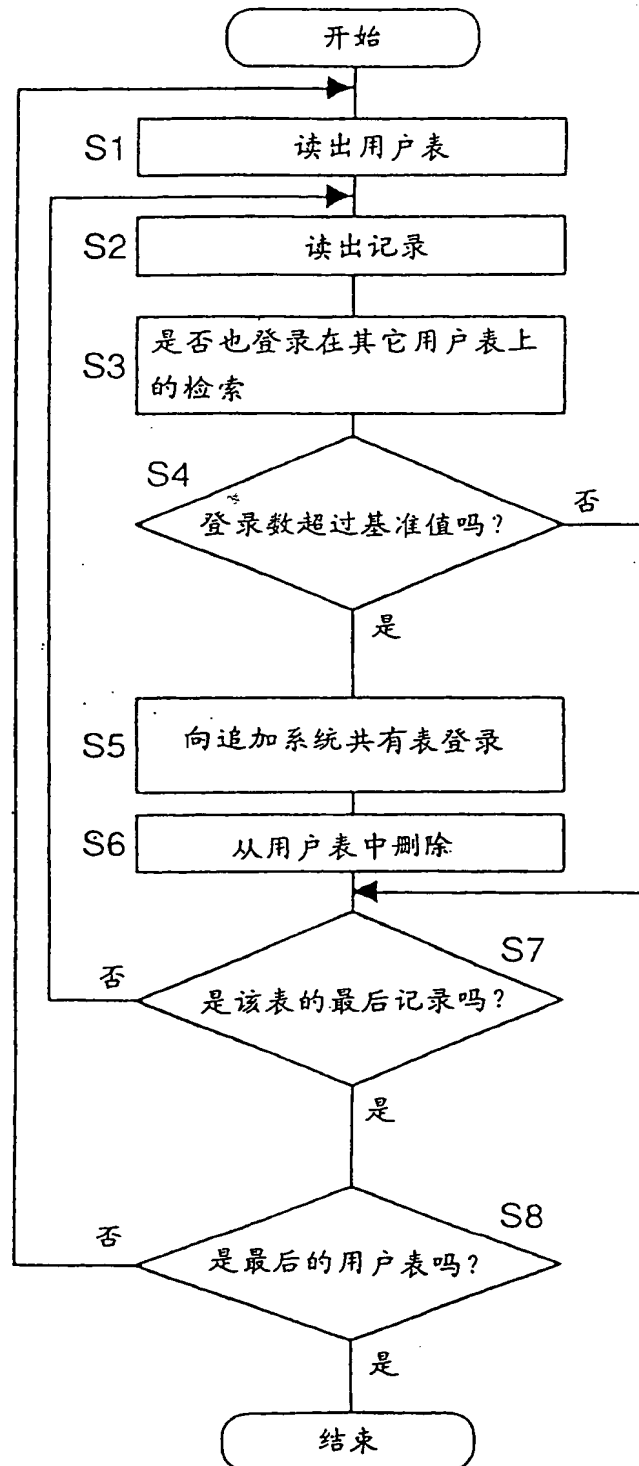


图 3

用户 1 的表

姓名	公司	公司电话	住宅电话
長岡宏	ナガオカヤ	075-957-9815	075-111-2222
田中一	タナカヤ	03-1111-2222	03-2222-3333
齊藤茂吉	第一電気	06-222-3333	06-333-4444
:	:	:	:

用户 2 的表

姓名	公司	公司电话	住宅电话
長岡宏	ナガオカヤ	075-957-9815	075-111-2222
山本太郎	ヤマモトヤ	06-123-4567	06-345-6789
鈴木花子	ハナコヤ	03-1234-5678	03-2345-6789
:	:	:	:

用户 N 的表

姓名	公司	公司电话	住宅电话
田中一	タナカヤ	03-1111-2222	03-2222-3333
夏目漱石	第一出版	03-4321-9876	03-5432-8765
島崎藤村	出版堂	06-9876-5432	06-7654-3210
:	:	:	:

追加系统共有表 (无)

姓名	公司	公司电话	住宅电话

图 4

用户 1 的表

姓名	公司	公司电话	住宅电话
齐藤茂吉	第一電気	06-222-3333	06-333-4444
⋮	⋮	⋮	⋮

用户 2 的表

姓名	公司	公司电话	住宅电话
山本太郎	ヤマモトヤ	06-123-4567	06-345-6789
鈴木花子	ハナコヤ	03-1234-5678	03-2345-6789
⋮	⋮	⋮	⋮

⋮

用户 N 的表

姓名	公司	公司电话	住宅电话
夏目漱石	第一出版	03-4321-9876	03-5432-8765
島崎藤村	出版堂	06-9876-5432	06-7654-3210
⋮	⋮	⋮	⋮

追加系统共有表(无)

姓名	公司	公司电话	住宅电话
長岡宏	ナガオカヤ	075-957-9815	075-111-2222
田中一	タナカヤ	03-1111-2222	03-2222-3333

图 5

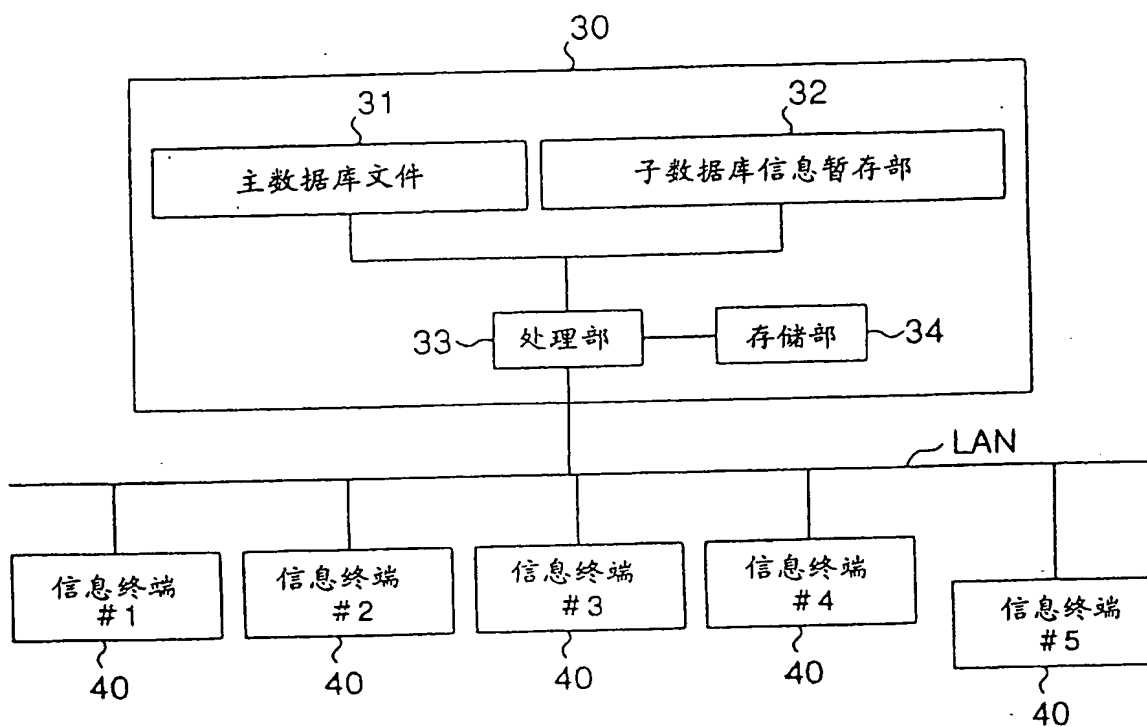
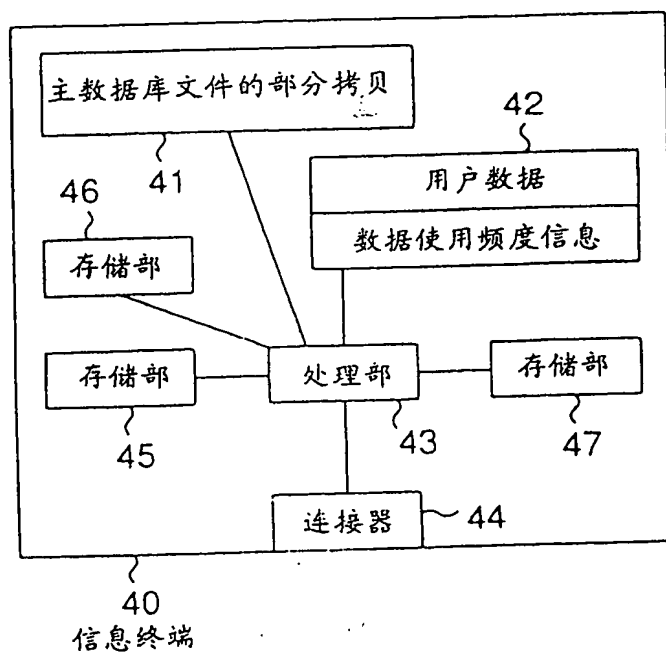


图 6



94-09-10

图 7

主数据库文件				
姓名	公司电话号码	住宅电话号码	使用频度	追加数据
山田花子	03-1212-3232	03-2132-9898	5	0
山下太郎	075-789-9876	075-543-9696	27	0
下田道子	06-7777-9595	06-2132-6461	20	0
大木聡美	033-234-3636	033-655-9632	55	0
⋮	⋮	⋮	⋮	⋮

图 8

信息终端 1 的子数据文件			
姓名	公司电话号码	住宅电话号码	使用频度
岡本太郎	075-333-3333	075-555-55555	30
齋田次郎	03-1234-4321	03-2222-2323	15
林三郎	06-5656-2323	06-7898-1111	5
⋮	⋮	⋮	⋮

信息终端 2 的子数据文件			
姓名	公司电话号码	住宅电话号码	使用频度
齋田次郎	03-1234-4321	03-2222-2323	10
岡本太郎	075-333-3333	075-555-55555	25
林三郎	06-5656-2323	06-7898-1111	30
久野四郎	033-233-3232	033-656-3232	30
⋮	⋮	⋮	⋮

信息终端 3 的子数据文件			
姓名	公司电话号码	住宅电话号码	使用频度
齋田次郎	03-1234-4321	03-2222-2323	5
岡本太郎	075-333-3333	075-555-55555	27
中野五郎	06-5656-1333	06-7898-6666	20
久野四郎	033-233-3232	033-656-3232	55
⋮	⋮	⋮	⋮

图 9

子数据库信息暂存部									
姓名	公司电话号码	住宅电话号码	使用者 名称 1	频度 1	使用者 名称 2	频度 2	使用者 名称 3	频度 3	频度 合计
岡本太郎	075-333-3333	075-555-5555	信息终端 1	30	信息终端 2	25	信息终端 3	27	82
林三郎	06-5656-2323	06-7898-1111	信息终端 2	30					30
久野四郎	033-233-3232	033-656-3232	信息终端 2	30	信息终端 2	55			85

图 10

主数据库文件					
姓名	公司电话号码	住宅电话号码	使用频度	追加数据	
山田花子	03-1212-3232	03-2132-9898	5	0	
山下太郎	075-789-9876	075-543-9696	27	0	
下田道子	06-7777-9595	06-2132-6461	20	0	
大木聡美	033-234-3636	033-655-9632	55	0	
岡本太郎	075-333-3333	075-555-5555	27	1	
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮

图 11

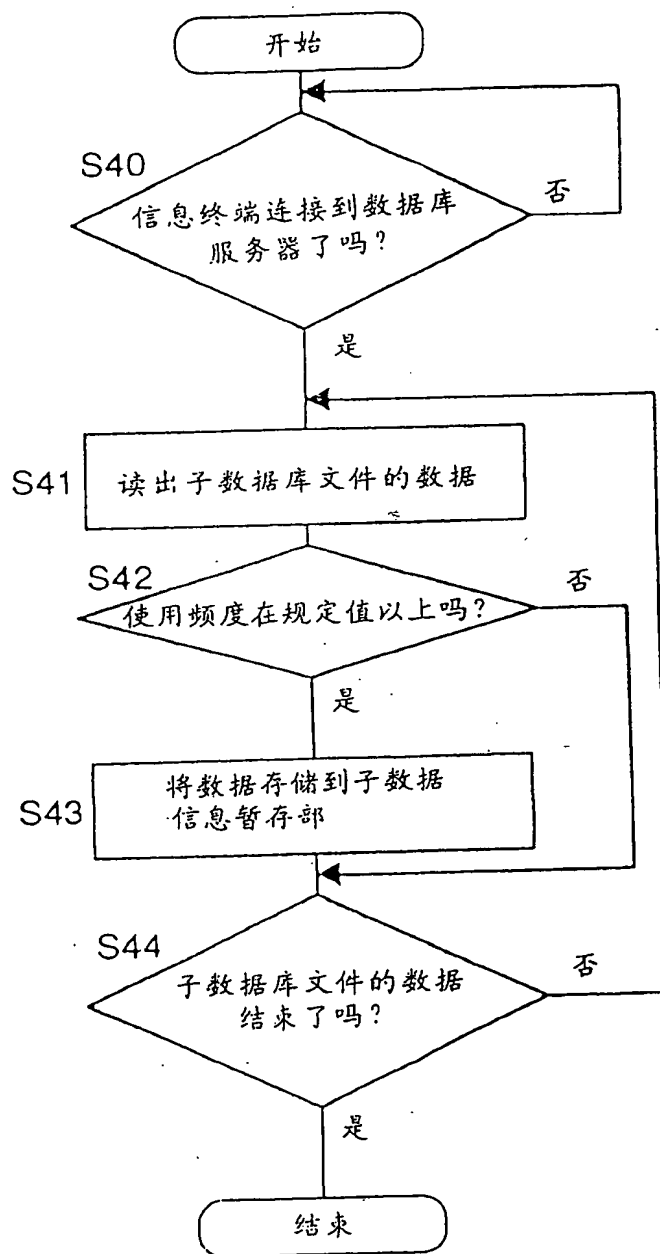


图 12

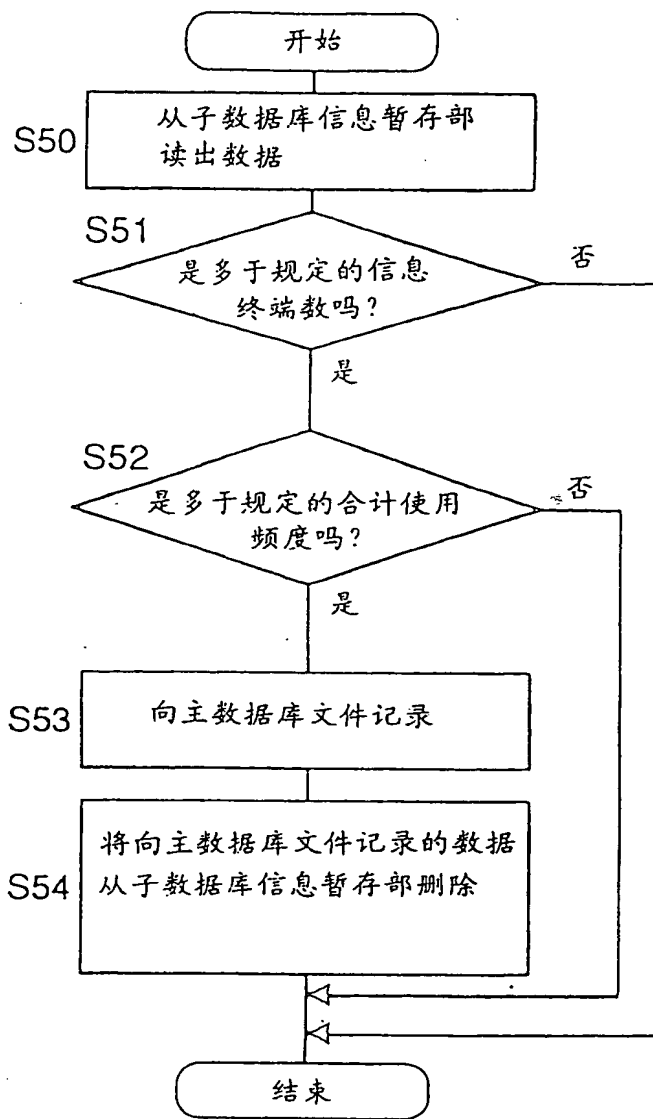


图 13

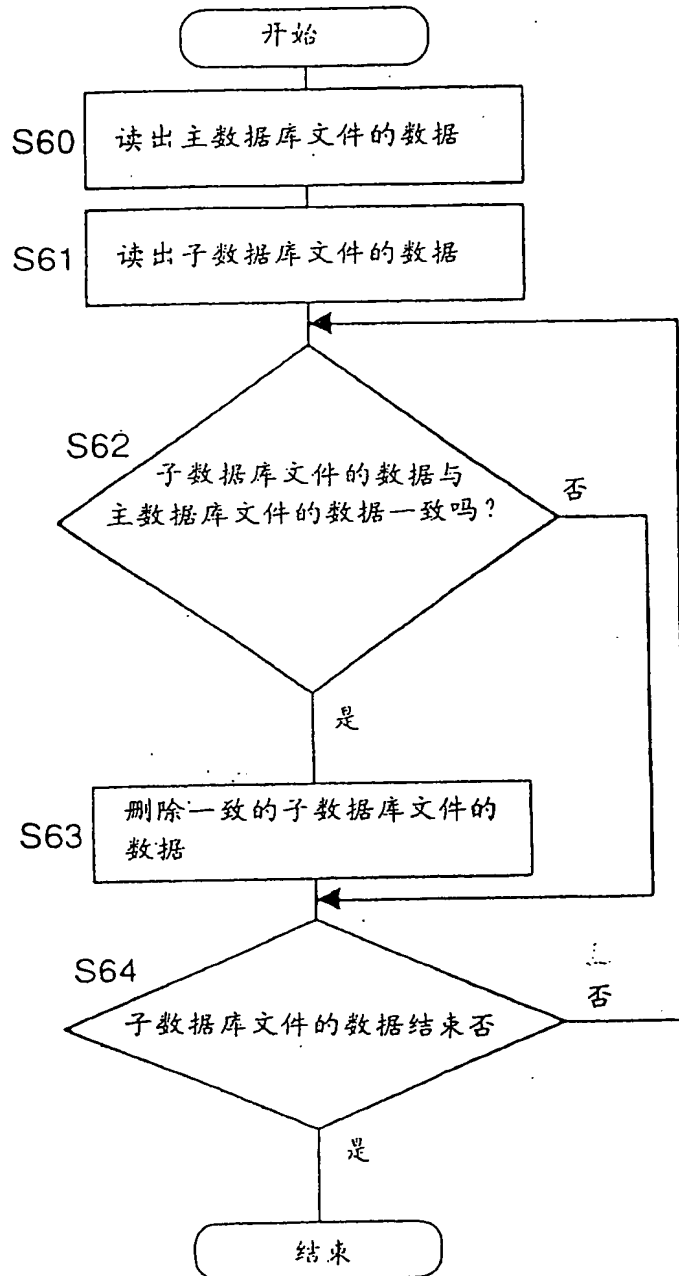


图 14

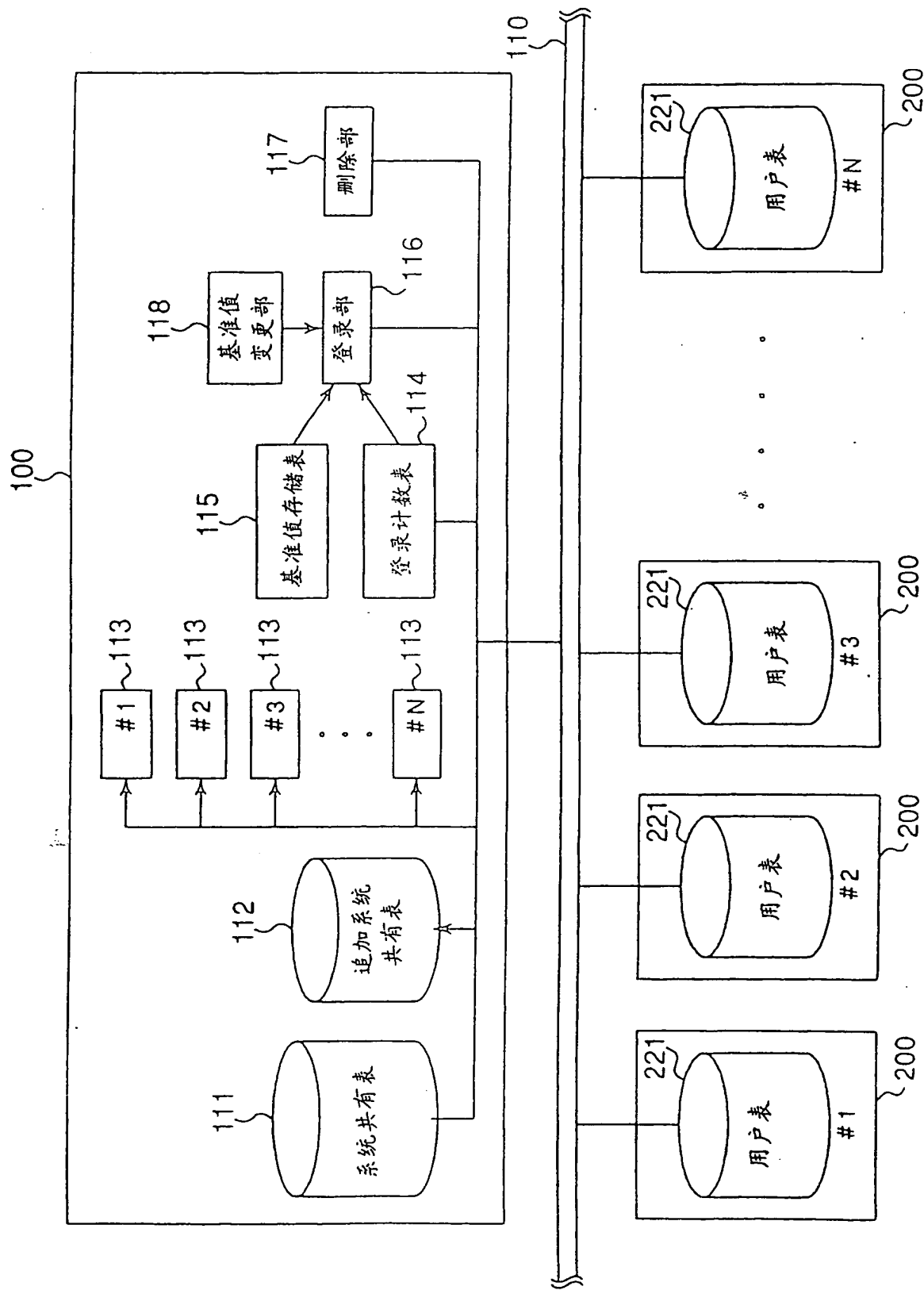


图 15

(固有名词)

141 读音	142 变换语	143 用户辞典数	144 合计值
ひろの	広野	5	10
くぼ	窪	10	15
.	.	.	.
.	.	.	.
.	.	.	.

(普通名词)

141 读音	142 变换语	143 用户辞典数	144 合计值
.	.	.	.
.	.	.	.
.	.	.	.

(动词)

141 读音	142 变换语	143 用户辞典数	144 合计值
.	.	.	.
.	.	.	.
.	.	.	.

图 16

151 词类	152 第 1 基准值 (用户辞典数)	153 第 2 基准值 (合计值)
固有名词	3	10
其它名词	5	15
动词	10	20
.	.	.
.	.	.
.	.	.

图 17

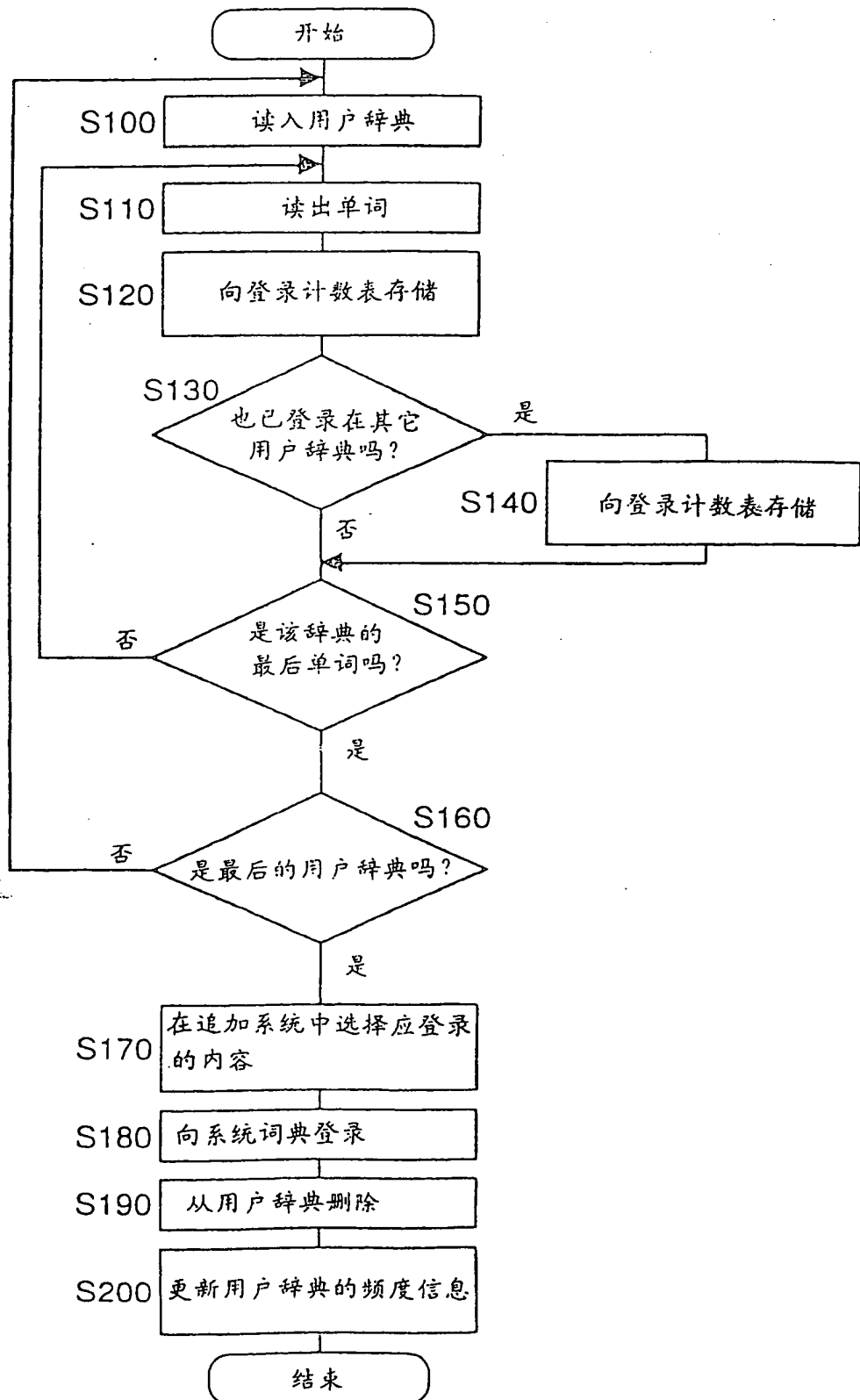
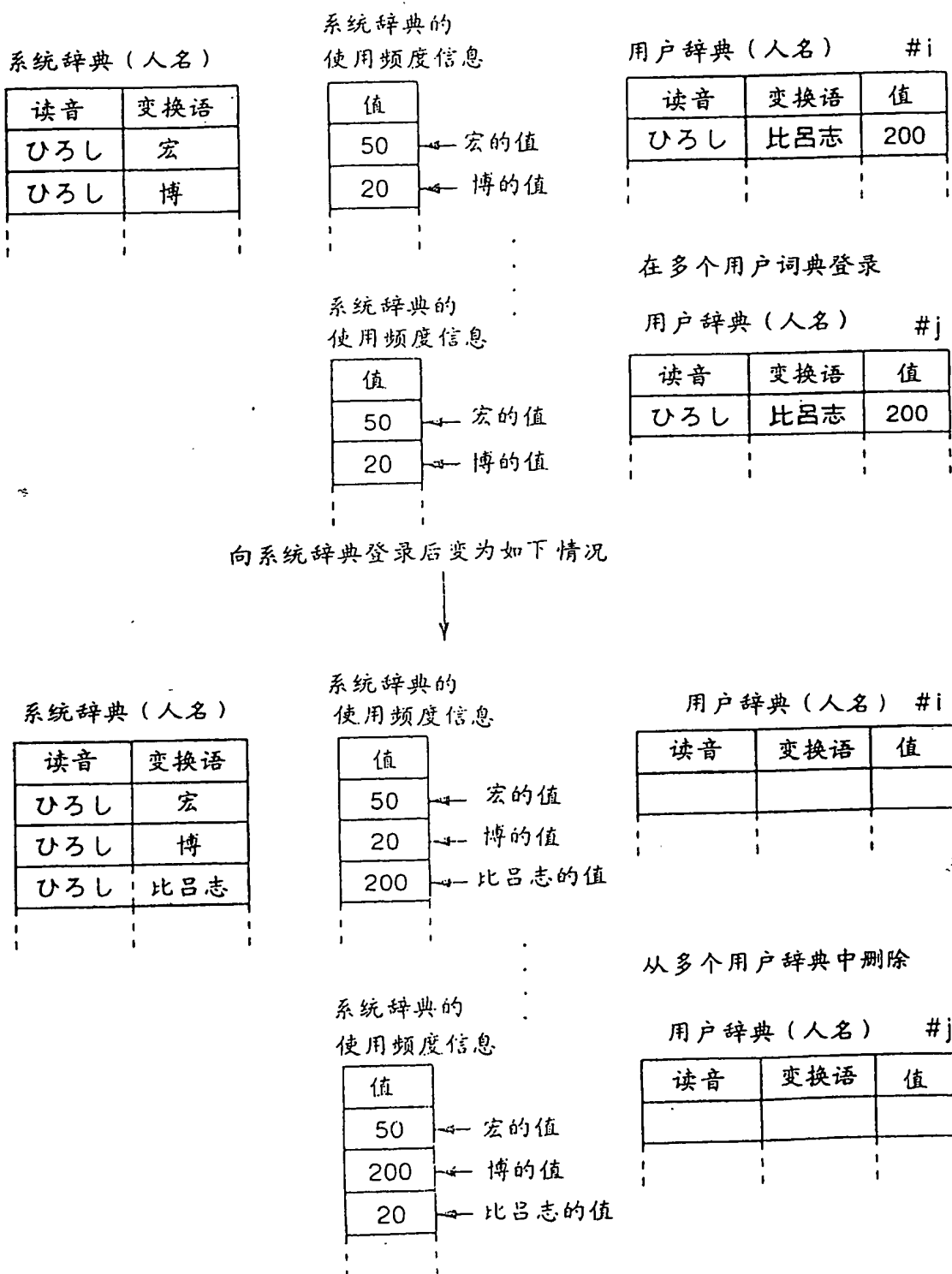


图 18



2004 10

图 19

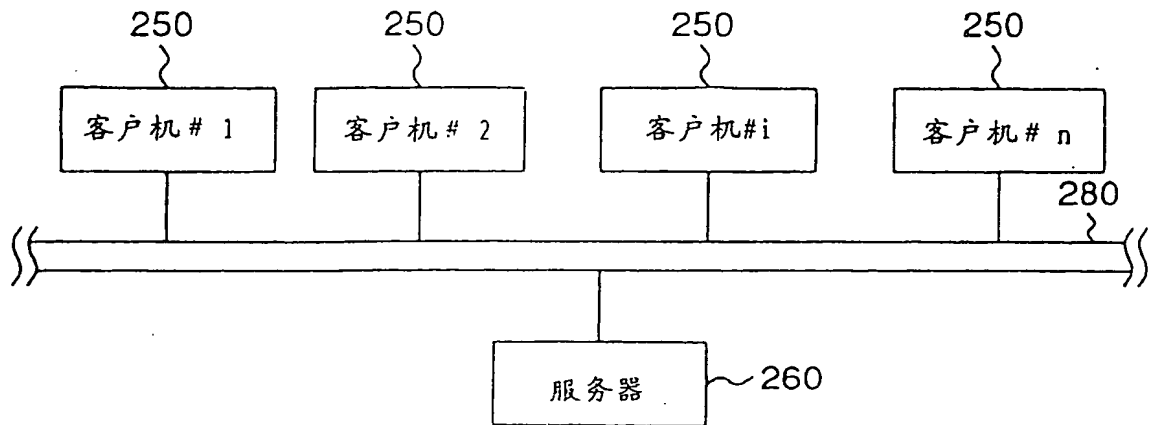


图 20

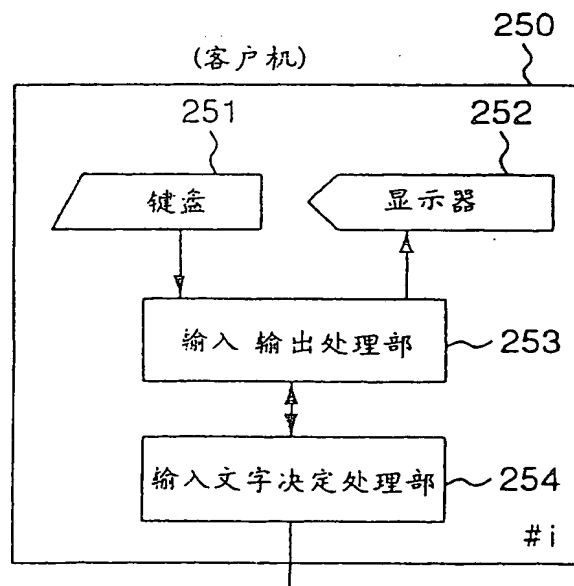


图 21

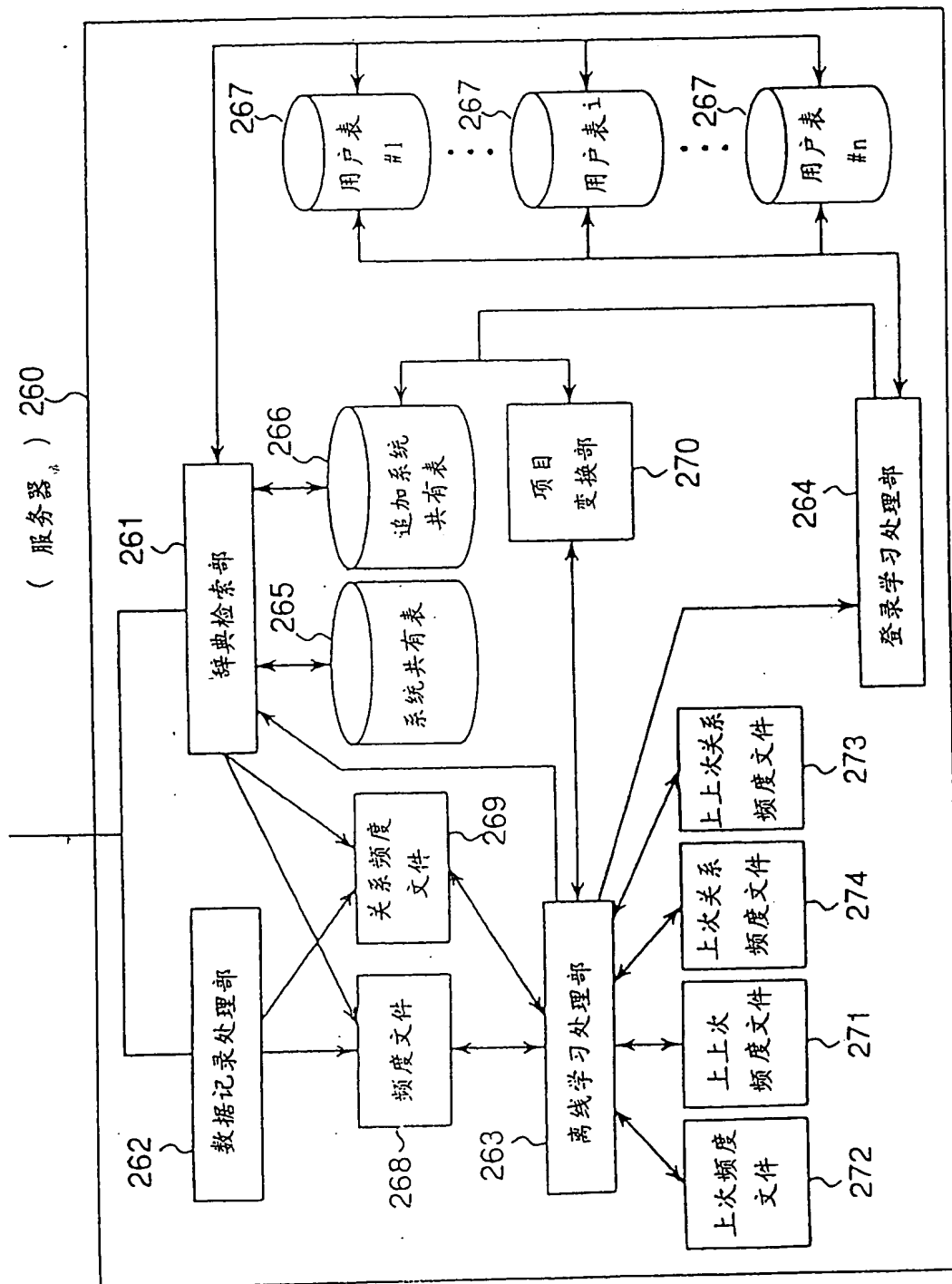


图 22

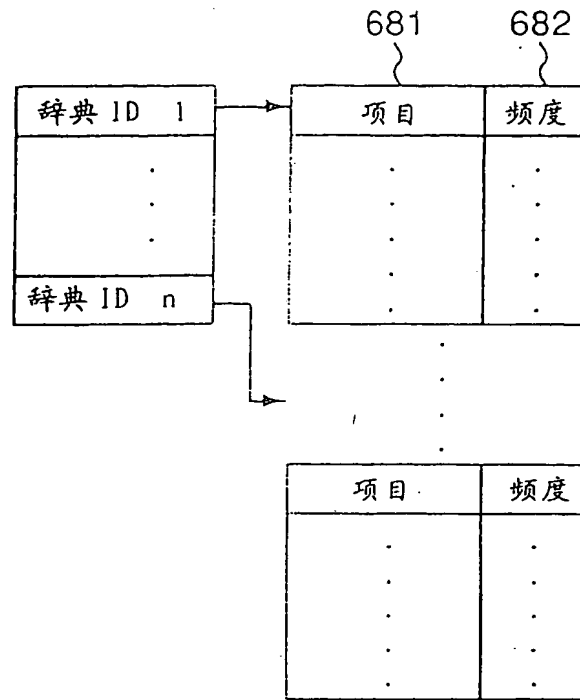


图 23

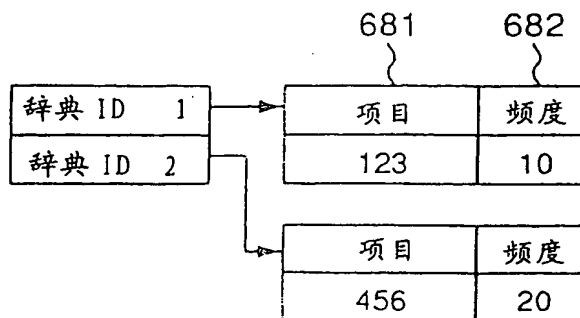
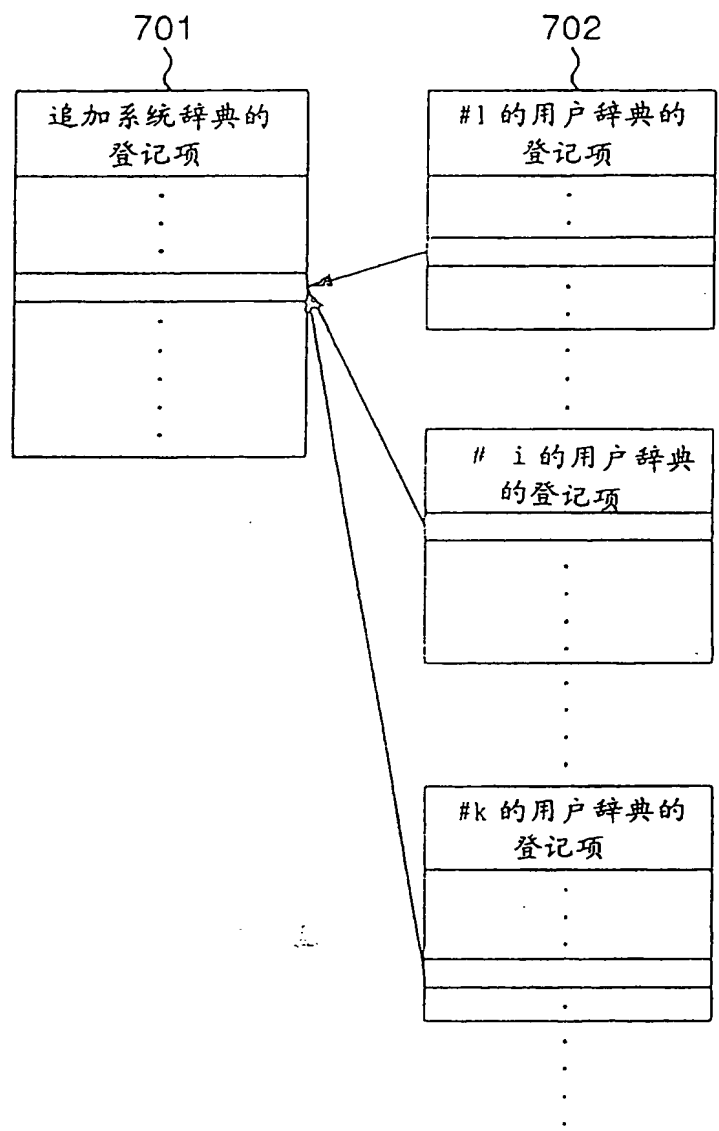




图 26



**Abstract**



图 28

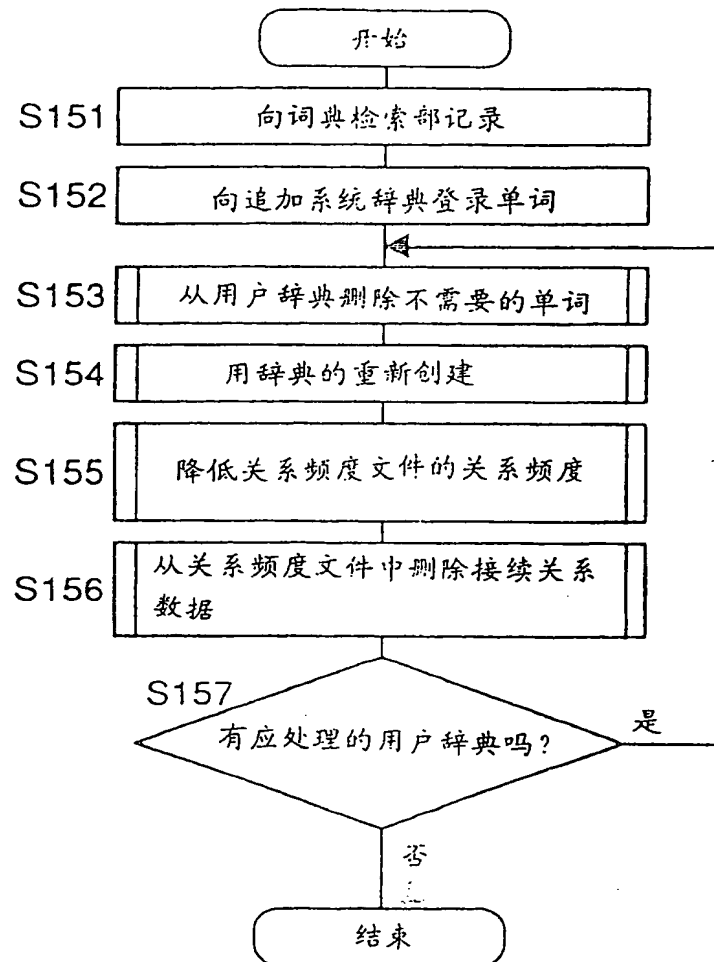


图 29

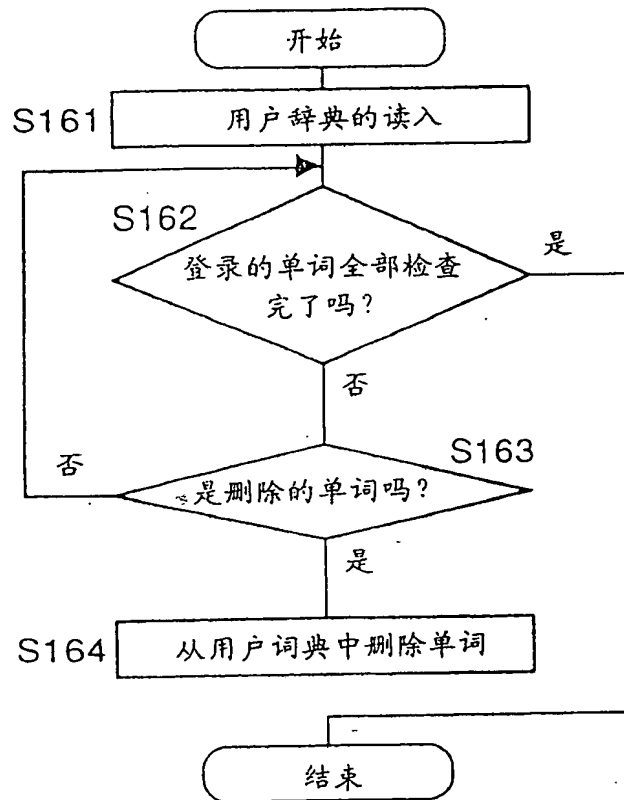


图 30

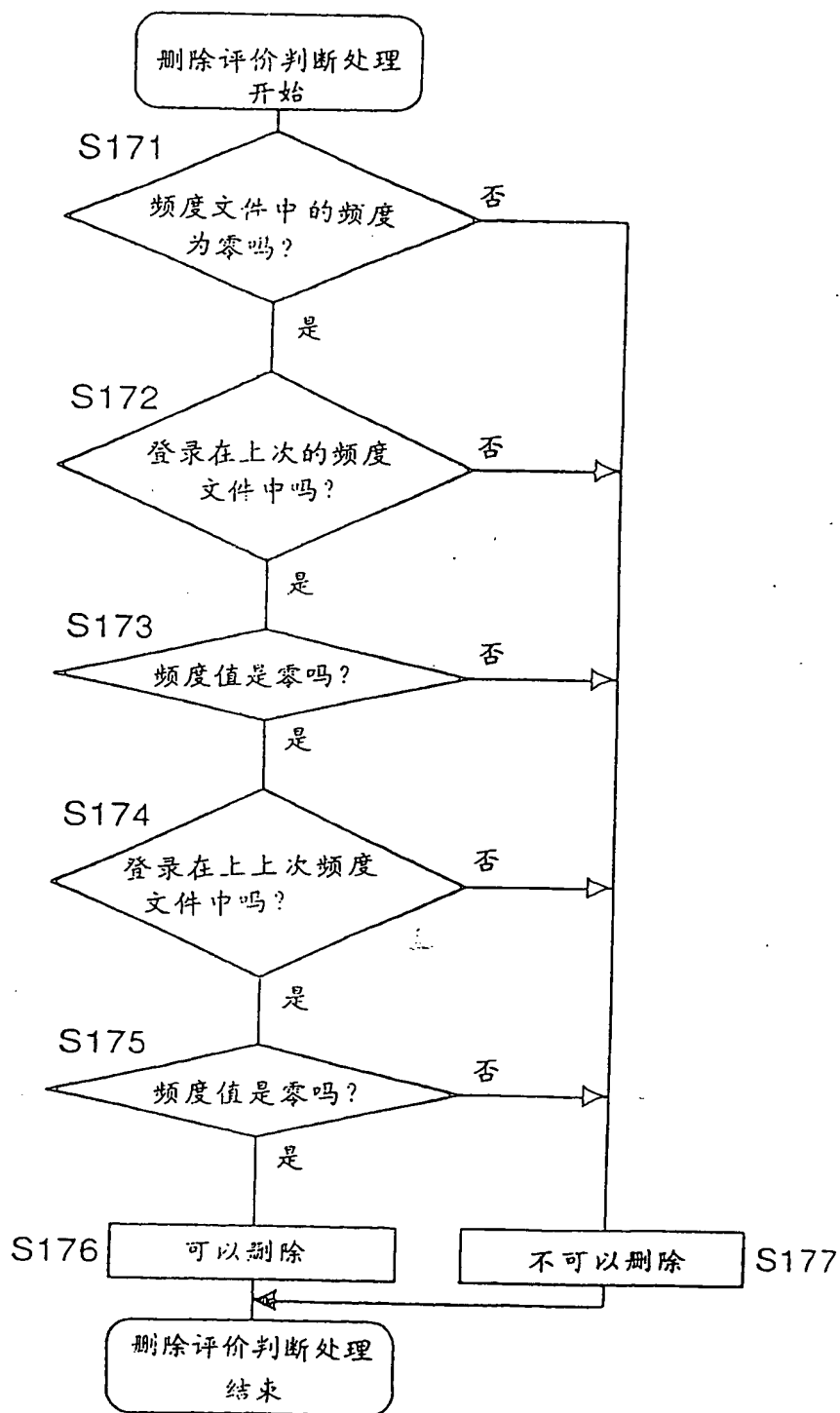


图 31

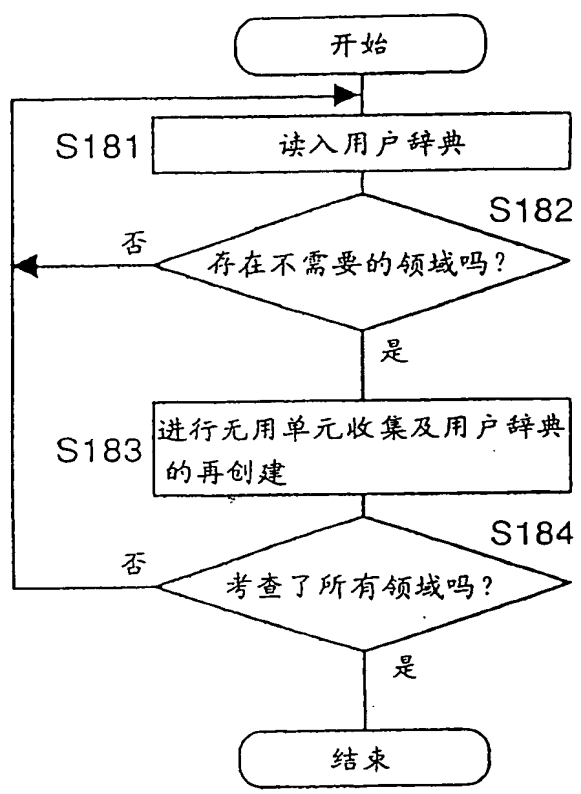


图 32

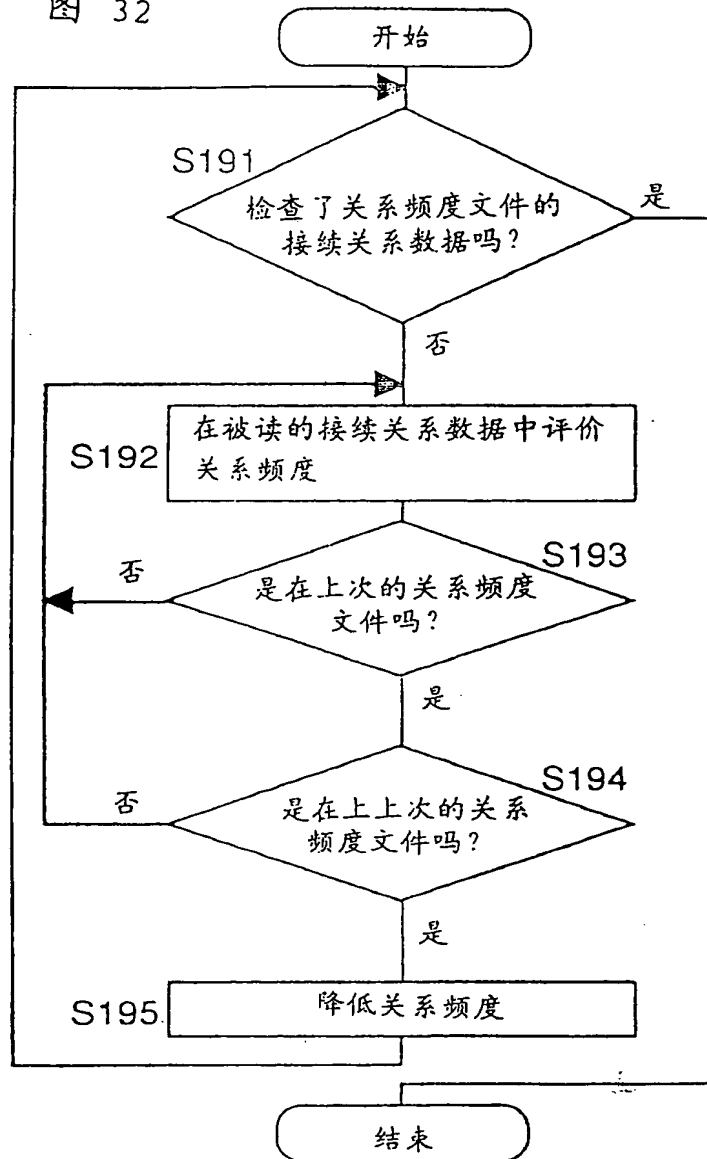
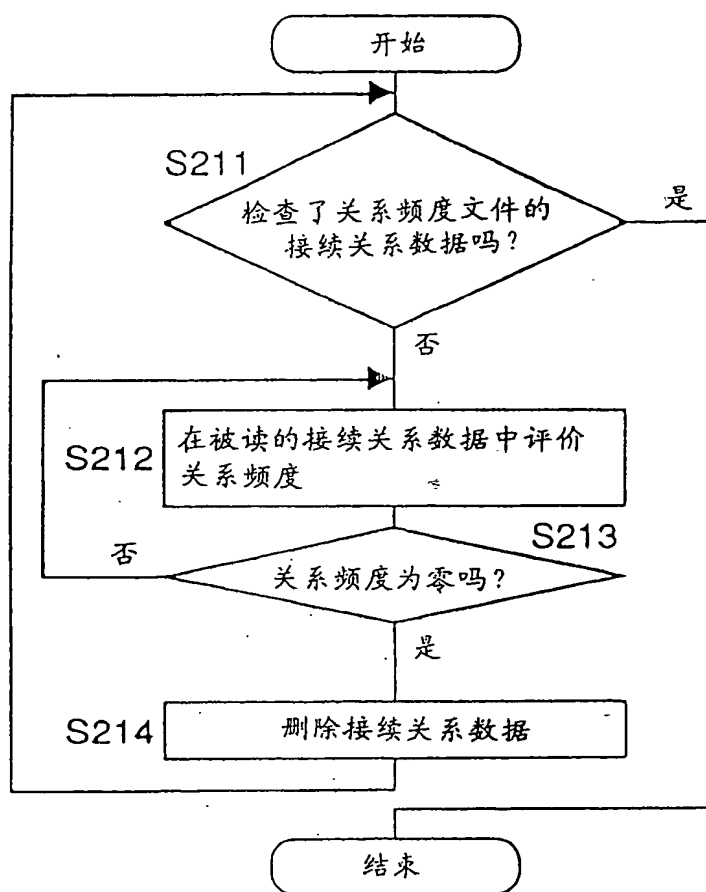


图 33

	上上次	上次	本次	降低关系 频度处理	
样板 1	• —→ • —→ •			处理	箭头表示关系频度的 趋向 —→ 不变 ↘ 关系频度下降
样板 2	• ↘ • —→ •			处理	
样板 3	样板 1、2 之外			不处理	

图 34



**This Page Blank (uspto)**